

[illegible]

RESULT 2

```

US-09-565-910-2
; Sequence 2, Application US/09565910
; Patent No. 6268192
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
; TITLE OF INVENTION: HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.

```

Db	660	GGCTTCTGCACTGAGTGGAAAGAGACCTCATCGATGCCCTACGGGGTCTAGTGGCAGC	719
Qy	567	GGCGTGGCTTCGTGTGTGCATTCTCCGAGGCCCTGGCTGAAGGAGCCGTCGAAGATGGC	626
Db	720	GGGCTCGCCTATGCATTCATGGCTCTGGAGGCATTGGCTGATGGTGGGGTGAAGATGGT	779
Qy	627	ATGCCAGAGGCTGGGCCACCGCATCGCTGCCAGACCCCTGTGGGGACGCCAAGATG	686
Db	780	TTGCGCACGCGCGCTGGCAATCCAATCGGGGCCAGGCTTTGCTGGGAGCTGCCAAGATG	839
Qy	687	CTGCTCCAGGAGGGCCACACCCAGCCAGCTGGCTCAGAGCTGTGCACCCCGGGTGGC	746
Db	840	CTGCTGGACTCGGAGCAGCATCCATGCCAGCTTTAAGGACAAATGTCTGCTCCCTCGGGGA	899
Qy	747	ACCACCATCTATGGACTCCAGCCCTCGAGCGGGGGGCTGGAGCACCACCATGAGC	806
Db	900	GCCACCATCCAGCCCTGCATCTTCAGAGAGTGGGGCTTCGCTCTCTGTCATCAAT	959
Qy	807	GGCGTGGAGGCTGCCAAGCTGCCGGGCCAAGGAGCT	841
Db	960	GCAGTTGAGGCTCTGTATCCGAACACGAGAGCT	994

RESULT 3

```

US-08-961-527-22
; Sequence 22, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

```

Query Match 5.7%; Score 75.8; DB 4; Length 28171;
Best Local Similarity 45.9%; Pred. No. 1.1e-08;
Matches 260; Conservative 0; Mismatches 307; Indels 0; Gaps 0;

QY 319 TCCTTGGTGTCGCTCGCTGGGTGCTCTCTCAGCACCCCTGGAGGAGCGTGTGCCCCCAA 378
||||| ||| ||||| ||||| ||| ||| ||| |||
Db 7365 TTTTGATTTTCGATGGGCAGCTGGATTGACCTTAGAAAACTAGCAAGTCTTATCCCAAGTC 7424

Qy	379	ACACACGGGTGCTCGGGTCTTGCCCAACCTGCCCTGTGTGGTCCAGGAAGAGGGCCATAG	438
Db	7425	AACACCGAAATTAATTCGTATGATGCTAATACCCCTGCTTCATCGGCAAGAGTGATTA	7484
Qy	439	TGATGGCGCGGGCGCCACGTGGGGAGCAGCAGACCAAGCTCCTGCCAGCATCTGCTGG	498
Db	7485	GTTATGCTTGTCTCCTAATTGACGGCTGAGACAGTGAGCTCTTTTATCAGCTTTTAG	7544
Qy	499	AGGCCTGTGGCGGTGTGAGGAGGTGCTTGAAGCCTACGTCGACATCCACACTGGCCTCA	558
Db	7545	CCAAGGCTGGTCTCTTGTTGAAC TAGGACAAAGATTAAATCGATGCAGCGACAGGCTTG	7604
Qy	559	GTGCAGTGCGTGGCGCTCGTGTGTCATCTCCGAGGCCCTGGCTGAGGAGCCGTCA	618
Db	7605	CAGTTGTGACACACCTTTGTCTATCTTTTATCGAGGCTTGGCAGATGCAAGTGTTC	7664
Qy	619	AGATGGCATGCCACAGCGCTGGCCACCGCATCTGCCACAGCCCTGCTGGGACGG	678
Db	7665	AGACAGATTACCACGAGAATAAGCATTTGAAATGGCAGCACAACTGTTGAGGAGCTG	7724
Qy	679	CCAAGATGCTGCTGCAGGAGGGCCACACCCAGCCAGCTGCCTCAGACGTGTGCACCC	738
Db	7725	GGCAATTTGCTTGAAGTCACACATCCTCGAGTATTGAAGACCAAGTCTGTAGCC	7784
Qy	739	CGGTGGCACCCATCTATGGACTCCAGCCCTGGAGCAGGGCGGCTCGCAGCAGCCA	798
Db	7785	CAGCGGTTCCGACTATCGCTGGTGTAGCAAGCCTAGAAGCGCATGCTTTCGAGGAACAG	7844
Qy	799	CCATGAGCCGCTGGAGGCTGCCACCTGCCGGCCAGGAGCTCAGCAAACTAGGCTG	858
Db	7845	TCATGGATGAGTTCATCAAGCCTACAAAGAACACAAAGAACTAGGTAATAAGAGGTAG	7904
Qy	859	GGCTCTGGCCATCCTTTCTCGCCTCTG	885
Db	7905	TTTTTGACTGCCTCTTTTATGGTGGCTG	7931

RESULT 4

```

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: FRASER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

Query Match	5.58;	Score 73.4;	DB 4;	Length 4403765;
Best Local Similarity	48.7%;	Pred. No. 1.6e-07;		
Matches 200;	Conservative	0;	Mismatches 211;	Indels 0; Gaps 0;

[illegible]

Db 591763 CACGATCGCGTATTTCGAATCCAAAGCTACCGCGGGACGCCAGTGGTGGCTGCATGCC 591822
QY 404 CAACTGCGCTGTGTGTCAGGAAGGGCCATAGTATGATGCGGGGCGCCACAGTGGG 463
Db 591823 GAACGCGCGGCGATTGGTGGAGCGGGGTACAGCGTGGCCAAAGCGCGCTTTGTAC 591882
QY 464 GAGCAGCGAGCAACAGCTCCTGCAGCATCTGCTGAGGCGCTGTGGCGGTGTGAGGAGGT 523
Db 591883 CCCGCAACAGCTTGAGGAGGTCTCGGCCCTTGTTCACCGGTGCGGCGGTGCTGACCGT 591942
QY 524 GCGTGAAGCCATAGTCGACATCCACACTGGCCCTAGTGGCAGTGGCGGTGGCGCTTCGTGTG 583
Db 591943 TCCGGAATCGCAGTTGGACGCGGTGACCGCGGTGTCCGCGTCCGGCTATTTCCTT 592002
QY 584 TGCATTCTCCGAGGCGCTGGCTGAAGGAGCGCTCAAGATGGCATGGCCACGAGCGCTGGC 643
Db 592003 TCTGCTGGTCGAGGCGCTGGTGGATGCGGAGTGGCGGTGGCTGGCTGACCGCTCAGGTGGC 592062
QY 644 CCACCGCATCGTGGCCAGACCCCTGCTGGGGACGCCAAGATGCTGCTGCA 694
Db 592063 CACCGATCTCGCGCGCAGACAATGGTGGCTCAGCGCGATGCTGCTGGA 592113

RESULT 5
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 5.5%; Score 73.4; DB 4; Length 4411529;
Best Local Similarity 48.7%; Pred. No. 1.6e-07;
Matches 200; Conservative 0; Mismatches 211; Indels 0; Gaps 0;
QY 284 GGCAGAGGTGCTTCCTGTGTCACCACTGAACACATCTTGTGTGTCGGTGGCTGGGGT 343
Db 590341 GACTGCGCGCGCGAAGAACACAGTCTGAGCAGGTGTTGCTCACCGTGGTAGCGGCAT 590400
QY 344 GTCTGTGAGCACCTGGAGAGGTGTCGCCCAACACACAGCGGTGCTGGGGTCTTGGC 403
Db 590401 CACGATCGCGTATTTCGAATCCAAAGCTACCGGTGGGACGCGAGTGGTGGCATGCC 590460
QY 404 CAACCTGCGCTGTGTGTCAGGAAGGGCCATAGTATGTCGCGGGCGCGCCACGCTGGG 463
Db 590461 GAACGCGGGCGATTGGTGGGACGGGTTACAGCGCTGGCCAAAGCGCGTTGTAC 590520
QY 464 GAGCAGCGAGACCAAGCTCCTGAGCATCTGTGGAGGCTGTGGCGGTGTGAGAGGT 523
Db 590521 CCCGCAACAGCTTGAGGAGTCTCGGCCCTTGTTCGACGCGGTGCGCGCGGTGCTGACCGT 590580
QY 524 GCCTGAGCCCTAGCTCGACATCCACACTGGCCCTCAGTGCAGTGGCGTGGCTTGTGTG 583
Db 590581 TCCGGATCGCAGTTGAGCGCGGTGACCGCGGTGTCCGGCTCGGGTCCGGGCTATTTCCT 590640
QY 584 TGCATTCTCGAGGCGCTGCTGAAGGAGCGCGCTCAAGATGGGATGGCCAGCAGCGCTGGC 643

Db 590641 TCTGCTGTGTCGAGGCGCTGCTGGATCCGAGTTCGGGGTGGCTTGAAGCGCTCAGGTGGC 590700
QY 644 CCACCGCATCGCTGCCAGACCCCTGCTGGGGACGGCCAAAGATGCTGCTGCA 694
Db 590701 CACCGATCTCGCGCGCAGACAATGGTGGCTCAGCGCGATGCTGCTGGA 590751
RESULT 6
US-08-665-716-1
; Sequence 1, Application US/08665716
; Patent No. 5789222
; GENERAL INFORMATION:
; APPLICANT: KELLY, ROSEMARIE
; APPLICANT: REGISTER, ELIZABETH A
; APPLICANT: MASUREKAR, PRAKASH S
; TITLE OF INVENTION: P5C REDUCTASE GENE FROM ZALERION
; TITLE OF INVENTION: ARBORICOLA
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MERCK & CO., INC.
; STREET: 126 E. LINCOLN AVENUE
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: US
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/665,716
; APPLICATION NUMBER: US/08/665,716
; FILING DATE: 23-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: KORSER, ELLIOTT
; REGISTRATION NUMBER: 32,705
; REFERENCE/DOCKET NUMBER: 19453PV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-594-5493
; TELEFAX: 908-594-4720
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1147 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..960
US-08-665-716-1

Query Match 5.0%; Score 66.6; DB 1; Length 1147;
Best Local Similarity 52.3%; Pred. No. 6.6e-07;
Matches 172; Conservative 0; Mismatches 154; Indels 3; Gaps 1;
QY 520 AGTGTGCTGAAGCGCTAGTTCGACATCCACACTGCGCTCAGTGGCAGTGGCGCTGCG 579
Db 651 AGCTCCCGCTGCTACACATGATGATGCCAGCAGCGGCTGTGTGGCAGTGGCCAGCTTTT 710
QY 580 TGTGTGCAATTCCTCGAGGCGCTGCTGAAGAGCGCTCAAGATGGGATGCCAGCAGCC 639
Db 711 TTGCGCTCATACTGAGAGCTCGGATTGATGGGCTGTGGCGATGGGTGCGCAAGACAG 770
QY 640 TGGCCCGACCGCATCGCTGCCAGACCCCTGCTGAGGAGCGCTCAAGATGCTGCTGCACGAGG 699
Db 771 AGGTCAGAGATGGCGCGCAGACTATGAAGGGGCTGCTGGATTGTTGTTCTGGAG 830
QY 700 GCCAACACCCAGCCAGCTGCGCTCAGACGTGTGCACCCCGGGTGGCACCACCATCTATG 759

Db 831 ---AGCATCGCGGTTGCTGAAAGATAAAGTGACTACGCGCGGTGGTCCACATAGGTG 887
QY 760 GACTCCACGCCCTGGAGCAGGCGGCTCGGAGCAGCCACCATGAGCGCGCTGGAGGTG 819
Db 888 GGTGATGTTGCTGGAAGAGGAGGCGGTGAGAGGACCGTGGCTAGAGCAGTTAGGGAAG 947
QY 820 CCACCTGCGCGCGGCCAGGAGGCTCAGCAGA 848
Db 948 CTACTGTGTTGCCAGTCAGCTGGTAAA 976

RESULT 7

US-09-154-083-17
; Sequence 17, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; FILE REFERENCE: Constructs Therefor
; FILE REFERENCE: 30062-2002.00 Polyketide Synthase Enz
; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-09-154-083-17

Query Match 3.2%; Score 42.8; DB 3; Length 466;
Best Local Similarity 46.4%; Pred. No. 0.2;
Matches 140; Conservative 0; Mismatches 162; Indels 0; Gaps 0;
QY 497 GGAGCGCTTGGCGGTGTGAGAGGTGCCTGAAGCCTACGTGACATCCACATGCGCT 556
Db 42 GGAGCGCGCGCGCGCGCGGTGAAGCGCTGCGGTGACGCCACGCTTCCACTCCGCGCT 101
QY 557 CAGTGGCAGTGGCGGCTTGTGTGTCATTCTCCGAGGCGCTTGAAGAGCGGT 616
Db 102 CATGACGCGATGCTCGAGCGCTTCCGCCAGTCCCGAAGCGGTGCTAGAGCGCC 161
QY 617 CAGATGGCATGCCACAGAGCGCTGGCCACCGCATCGTCCACAGACCCCTGCTGGGAC 676
Db 162 GCGGATCGGATCGTCTCGACCTGACCGCGCGTCTCGCGCGCGGCGAGATCGGCAC 221
QY 677 GCGCAAGATGCTGTCAGAGGCGCAACACCCAGCTGCGCTCAGAGTGTGCAC 736
Db 222 GCGCCACTACTGGGTGGCGCGCTCGCCACCTTCTGTGAGCTGGCGCGGCTGCGAGCG 281
QY 737 CCGGGTGGCACCACTATGAGTCCAGCCCTGGAGCAGGCGGCGCTGCGAGCAGC 796
Db 282 ACTGGAGCGCCAGGCGCTCGCCACCTTCTGTGAGCTGGCGCGGCGGCTGCTCGCG 341
QY 797 CA 798
Db 342 CA 343

RESULT 8

US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria

; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 3.2%; Score 42.2; DB 1; Length 7218;
Best Local Similarity 12.4%; Pred. No. 0.59;
Matches 17; Conservative 81; Mismatches 39; Indels 0; Gaps 0;
QY 837 GAGCTCAGCAAGTAGGTGGCTCTGGCCATCTCTGCTCTGCTGCTGCTGCTGCT 896
Db 1035 GAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1094
QY 897 CTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 956
Db 1095 YY 1154
QY 957 CTGCT 973
Db 1155 YYYYYYYYYYYYYYYYYY 1171

RESULT 9

US-09-249-585A-2/c
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2

Result No.	Query	Score	Query			DB	ID	Description
			Match	Length	DB			
1	613	44.2	279	2	T23765	hypothetical prote		
2	584.5	42.2	319	1	A41770	pyrroline-5-carbox		
3	502.5	36.3	273	2	T06477	probable pyrroline		
4	492	35.5	274	1	S10186	pyrroline-5-carbox		
5	480	34.6	270	2	G97299	pyrroline-5-carbox		
6	475	34.3	276	1	QJ2334	pyrroline-5-carb		
7	469	33.8	266	2	AE1124	1-pyrroline-5-carb		
8	458	33.0	266	2	AG1484	pyrroline-5-carbox		
9	455.5	32.9	270	2	AG1867	pyrroline-5-carbox		
10	423.5	30.6	269	1	RD8CC	pyrroline-5-carbox		
11	423.5	30.6	269	2	D90683	pyrroline-5-carbox		
12	423.5	30.3	269	2	H85533	pyrroline-5-carbox		
13	420.5	30.3	284	2	T36286	pyrroline-5-carbox		
14	419.5	30.3	284	2	T0549	pyrroline-5-carbox		
15	412.5	29.8	267	2	A76767	pyrroline-5-carbox		
16	395	28.5	265	2	H95107	pyrroline-5-carbox		
17	390.5	28.2	263	2	D71281	probable pyrroline		
18	387	27.9	261	2	JC2078	pyrroline-5-carbox		
19	385	27.8	265	2	B97976	probable pyrroline		
20	383.5	27.7	294	2	S72897	pyrroline-5-carbox		
21	366	26.4	295	2	G70745	pyrroline-5-carbox		
22	362	26.1	299	2	T29226	probable proc prot		
23	360.5	26.0	277	2	H90194	hypothetical prote		
24	354	25.5	270	2	E96982	pyrroline-5-carbox		
25	354	25.5	279	2	G93837	pyrroline-5-carbox		
26	349.5	25.2	264	2	C75385	pyrroline-5-carbox		
27	349	25.2	262	2	D86860	pyrroline-5-carbox		
28	345.5	24.9	273	2	JQ0418	pyrroline-5-carbox		
29	344	24.8	282	2	T50305	Delta 1-pyrroline-		

[illegible]

69 CLLVIFATKPHVLPVLAEEVAPVVTTEHILSVSVAAGVSLSTEELLPPNTRVLRVLPNLP 128

R.: Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloesch, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, D.; Jones, L.M.; Karst, U.

69 CLLVIFATKPHVLPVLAEEVAPVVTTEHILSVSVAAGVSLSTEELLPPNTRVLRVLPNLP 128

Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund,
A.; Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AG1484

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-266 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC95647.1; PID:g16412843; GSPDB:GN00178

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: lin0414

C:Superfamily: pyrroline-5-carboxylate reductase

Query Match 33.0%; Score 458; DB 2; Length 266;

Best Local Similarity 37.5%; Pred. No. 1.4e-28;

Matches 98; Conservative 58; Mismatches 97; Indels 8; Gaps 2;

QY 9 RVGFGVAGRWAGAIAGQLIRACKVEAOHILASAPTRNLCHFQAL-----GORTHSNQ 63

DB 2 KKIGFIGAGNAGMAMGLKANGLKEEDILVCG---RDEKLPKLTFFNGIQLTDTIA 58

QY 64 EVLQSCLLVIFARKPHVPAVLAEEVAPVVTTEHILVSVAGVSLSTLELLPPNTRVLRV 123

DB 59 QLABQADIIILSVKPTTPEILTAVKDKITPEKIVISVAGVTKIDLEELTSTETKIVRV 118

QY 124 LPNLPVQVQGAIVMARGHRVGSSETKLLOHLLLEACGRCEVEPEAYVDIHTGLSGSVAF 183

DB 119 MPNTPALVGVAMSSISPNSTAEETAHTIFSSFGKAEVVAENLMDAVGVGSGSPAY 178

QY 184 VCAFSEALBAGVAKMGMPSSLAHRIAQAOTLLGTAKMLLHGGHQAQLRSDVCTPGGTTTY 243

DB 179 VYMFIEALADGVLKMPDKAYKFAQAQVLAAGKAVLGTGHPGKLKMDVTPSGGTTIE 238

QY 244 GLHLEOGLRAATMSAVEAA 264

DB 239 AVKSLEDTGFRSSVISAVQAA 259

RESULT 9

AG1867

pyrroline-5-carboxylate reductase [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AG1867

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AG1867

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-270 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA072446.1; PID:g17129833; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr0488

C:Superfamily: pyrroline-5-carboxylate reductase

Query Match 32.9%; Score 455.5; DB 2; Length 270;

Best Local Similarity 40.9%; Pred. No. 2.2e-28;

Matches 110; Conservative 43; Mismatches 109; Indels 7; Gaps 6;

QY 10 RVGFGVAGRWAGAIAGQLIRACKVEAOHILASAP--TDNRLCHFQALGCRTHSNQEVLS 68

DB 4 KFLILGGVGMGALLSLRIARGIYQPSVIVSEPTARQAFLOQKYHGVVTTDNLVFTQ 63

QY 69 CLLVTF-ATKPHVLAEEVAPVVTTEH--ILVSVAGVSLSTLELLPPNTRVLRVLP 125

DB 64 AQDVVFLAVKPOVSAIAQELADVTDTDSPLVSVILAGVSLSQLEAAF--PQSPVIRMP 122

QY 126 NLPVQVQGAIVMARGHRVGSSETKLLOHLLLEACGRCEVEPEAYVDIHTGLSGSVAFVC 185

DB 123 NTPATVSGMTAICSGAYTAHQKLAQIFSAVGEVVEVSESLMDAVTGLSGSGPAYVA 182

QY 186 AFSEALBAGVAKMGMPSSLAHRIAQAOTLLGTAKMLLHGGHQAQLRSDVCTPGGTTIYG 244

DB 183 LLVEALADGGVAGSLPRGIANQLALQATVIGTAQ--LLHESLHPAELKDRVTSPPGGTTIAG 241

QY 245 LHLEOGLRAATMSAVEAAATCRAKELSR 273

DB 242 IAQLEKAGFRSALIEAVKAATWRSQELGK 270

RESULT 10

REDEC

pyrroline-5-carboxylate reductase (EC 1.5.1.2) - *Escherichia coli* (strain K-12)

C:Species: *Escherichia coli*

C:Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 01-Mar-2002

C:Accession: A00385; B64767

R:Deutch, A.H.; Smith, C.J.; Rushlow, K.E.; Kretschmer, P.J.

Nucleic Acids Res. 10, 7701-7714, 1982

A:Title: *Escherichia coli* delta(1)-pyrroline-5-carboxylate reductase: gene sequence,

A:Reference number: A00385; MUID:83116986; PMID:6296787

A:Accession: A00385

A:Molecule type: DNA

A:Residues: 1-269 <DEU>

A:Cross-references: GB:J01665; NID:g147358; PIDN:AAA86433.1; PID:g147359

A:Note: parts of this sequence, including the amino and carboxyl ends of the mature p

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: B64767

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-269 <BLAT>

A:Cross-references: GB:AE000145; GB:U00096; NID:g1786580; PIDN:AAC73489.1; PID:g17865

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: proC

A:Map position: 9 min

C:Function:

A:Description: catalyzes reduction of pyrroline-5-carboxylate to proline

A:Pathway: proline biosynthesis

A:Note: third enzyme in the proline biosynthetic pathway

C:Superfamily: pyrroline-5-carboxylate reductase

C:Keywords: oxidoreductase; proline biosynthesis

Query Match 30.6%; Score 423.5; DB 1; Length 269;

Best Local Similarity 34.6%; Pred. No. 7e-26;

Matches 92; Conservative 56; Mismatches 117; Indels 1; Gaps 1;

QY 9 RVGFGVAGRWAGAIAGQLIRACKVEAOHILASAPT--DRNLCHFQALGCRTHSNQEVLSQ 67

DB 3 KKIGFIGCGNMKAILGLLIASQVLPQGIWYVTPSPDKVAALHDPQFGINAESAQEAQ 62

QY 68 SCLLVIFATKPHVLAEEVAPVVTTEHILVSVAGVSLSTLELLPPNTRVLRVLPNL 127

DB 63 IADITFAAVKPGIMIKVLSEITSSLNKSLVSIAGVTLDOALALGHDRIIRAMPNT 122

QY 128 PCVQVQGAIVMARGHRVGSSETKLLOHLLLEACGRCEVEPEAYVDIHTGLSGSVAFVCA 187

DB 123 PALVNAAGMTSVTPNALVTPEADTLNIFRCFGEAEVIAEPMIHPVVGSGSPAYVFMF 182

QY 188 SEALBAGVAKMGMPSSLAHRIAQAOTLLGTAKMLLHGGHQAQLRSDVCTPGGTTIYGLHA 247

DB 183 IEAMADAAVLGMPRAQAYKFAAQAVMGSAKMWLETGEHPGALKDMVCSPPGGTTIEAVRV 242

QY 248 LEQGLRAATMSAVEAAATCRAKELSR 273

DB 243 LEEKGFRAAVIEAMTKMCKSEKLSK 268

RESULT 11

D90683
pyrroline-5-carboxylate reductase [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: D90683
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90683
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <HAY>
A:Cross-references: GB:BA000007; PID:BA033859.1; PID:g13359893; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs0437
C:Superfamily: pyrroline-5-carboxylate reductase

Query Match 30.6%; Score 423.5; DB 2; Length 269;
Best Local Similarity 34.6%; Pred. No. 7e-26;
Matches 92; Conservative 56; Mismatches 117; Indels 1; Gaps 1;
QY 9 RRVGFGAGRMAGTAOGLIRAGKVEAQHILASAPT-DRNLCHFQALGCRTHSHNOEVLQ 67
Db 3 KKGIFGCGNMGKAILGLIASGVLPQGIWYTPSPDKVAALHDKFGINAESAQEVAAQ 62
QY 68 SCLLVIFATKPHVLPVLAELAVPVVTEHILVSVAAAGVSLSTLEELLPPNTRVLRVLPNL 127
Db 63 IADIIFAAVKPGIMIKVLSEITSSLNKDSLVSIAAGVTLDQLARALGHDRIIRAMPNT 122
QY 128 PCVQVQGAIVMARGHVSSETKLQHLLEACGRCVEEPEAYVDIHTGLSGGVAFVCAF 187
Db 123 PALVNAGMTSVTPNALVTPEDADVLNIFRCFGEAEVIAEPMIHPVWVGSGSPAYVFMF 182
QY 188 SEALAEGAVKMGMPSSLAHRIAATAOGLIRAGKVEAQHILASAPT-DRNLCHFQALGCRTHSHNOEVLQ 247
Db 183 IEAMADAAYLGMCPRAQAYKFAQAQVMSKAWLETGEHPGALKDMVCSFGGTTIEAVRV 242
QY 248 LEQGLRAATMSAVEAATCRAKELSR 273
Db 243 LEEKGFRAAVIEAMTKCKMEKSEKLSK 268

RESULT 12

H85533
pyrroline-5-carboxylate reductase [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85533
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85533
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <STO>
A:Cross-references: GB:AE005174; NID:g12513226; PID:AA054732.1; GSPDB:GN00145; UWGP:Z04
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: PROC
C:Superfamily: pyrroline-5-carboxylate reductase

Query Match 30.6%; Score 423.5; DB 2; Length 269;
Best Local Similarity 34.6%; Pred. No. 7e-26;
Matches 92; Conservative 56; Mismatches 117; Indels 1; Gaps 1;

QY 9 RRVGFGAGRMAGTAOGLIRAGKVEAQHILASAPT-DRNLCHFQALGCRTHSHNOEVLQ 67
Db 3 KKGIFGCGNMGKAILGLIASGVLPQGIWYTPSPDKVAALHDKFGINAESAQEVAAQ 62
QY 68 SCLLVIFATKPHVLPVLAELAVPVVTEHILVSVAAAGVSLSTLEELLPPNTRVLRVLPNL 127
Db 63 IADIIFAAVKPGIMIKVLSEITSSLNKDSLVSIAAGVTLDQLARALGHDRIIRAMPNT 122
QY 128 PCVQVQGAIVMARGHVSSETKLQHLLEACGRCVEEPEAYVDIHTGLSGGVAFVCAF 187
Db 123 PALVNAGMTSVTPNALVTPEDADVLNIFRCFGEAEVIAEPMIHPVWVGSGSPAYVFMF 182
QY 188 SEALAEGAVKMGMPSSLAHRIAATAOGLIRAGKVEAQHILASAPT-DRNLCHFQALGCRTHSHNOEVLQ 247
Db 183 IEAMADAAYLGMCPRAQAYKFAQAQVMSKAWLETGEHPGALKDMVCSFGGTTIEAVRV 242
QY 248 LEQGLRAATMSAVEAATCRAKELSR 273
Db 243 LEEKGFRAAVIEAMTKCKMEKSEKLSK 268

RESULT 13

T36286
pyrroline-5-carboxylate reductase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T36286
R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21603
A:Accession: T36286
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-284 <SEE>
A:Cross-references: EMBL:AL049819; PID:CA042663.1; GSPDB:GN000070; SCOEDB:SCE7.04c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: proc; SCOEDB:SCE7.04c
C:Superfamily: pyrroline-5-carboxylate reductase

Query Match 30.3%; Score 420.5; DB 2; Length 284;
Best Local Similarity 35.4%; Pred. No. 1.3e-25;
Matches 95; Conservative 53; Mismatches 119; Indels 1; Gaps 1;

QY 5 EPSRRYGVGAGRMAGTAOGLIRAGKVEAQHILASAPT-DRNLCHFQALGCRTHSHNOE 64
Db 13 EPMTQKVAVLTGKIGELLSGMIGAGWAPAD-LVTARRRERADELRHGVTPVTNAE 71
QY 65 VLQSLVIFATKPHVLPVLAELAVPVVTEHILVSVAAAGVSLSTLEELLPPNTRVLRVL 124
Db 72 AAKAADTLITLVKPDQMTLDELAPHVPADRLVISGAAGVPTSFEEERLAPGTPVVRVM 131
QY 125 PNLPCVQVQGAIVMARGHVSSETKLQHLLEACGRCVEEPEAYVDIHTGLSGGVAFV 184
Db 132 TMTALVDEAMSVISAGTATTAHLLTHTTEIFGAVGKTLRVPESEQDQACTALSQSPAYF 191
QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAATAOGLIRAGKVEAQHILASAPT-DRNLCHFQALGCRTHSHNOE 244
Db 192 FYLYEAMTDAGILGLPRDKAHLIVQSAIAAKMLRDSGEHPVKLRNVTSPAGTTINA 251
QY 245 LHALEQGLRAATMSAVEAATCRAKELS 272
Db 252 IRELENHGVRAALIAALEAARDRSRELA 279

RESULT 14

AI0549
pyrroline-5-carboxylate reductase [imported] - Salmonella enterica subsp. enterica se
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AI0549
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A>Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AI0549
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <PAR>
A:Cross-References: GB:AL513382; PID:CAD08841.1; PID:gl6501655; GSPDB:GN00176
C:Genetics:
C:Superfamily: pyrroline-5-carboxylate reductase

Query Match	30.3%;	Score	419.5;	DB 2;	Length	269;			
Best Local Similarity	34.6%;	Pred. No.	1.4e-25;						
Matches	92;	Conservative	57;	Mismatches	116;	Indels	1;	Gaps	1;
QY	9	RRVFGVAGRMAGAAOGLIRACKVQAQHILASAPT-DENLICHFOALGCRTHHSNQEVQLQ	67						
DB	3	KKIGFICGNGRAILGGLTASQVLPFGQIWIYTPSPDKVAALHDQYQGINAAQSAQAEVAQ	62						
QY	68	SCLLVIFATPHVLVAFLAEPVWTEHLVSVAAQVSLSTLEELLPTNPRVLRVLPNL	127						
DB	63	VADIVFQAVPGIMVKVLSISSLNKDSLWVIAAGVTLQDARALGHDKRRIRAMPNT	122						
QY	128	PCVVOEQAIVMARGHRHVGSGSETKLLQHLLEACGRCEEPEAYVDIHTGLSGSGVAFVCAF	187						
DB	123	PSLVNAGMTSVTPNALVTPEDTADVNLNIFPCFGEAEVTAEPMIHPVGVGSSGSPAYVFMF	182						
QY	188	SEALAEGAVKMGMPSSLAHRIAAROTLLGTAKMLLHEGQHAPOLRSDVCTPGGTTIYGLHA	247						
DB	183	TEAMADAALVGGMPRAQYAFKAAQVMSAKIVLETGKHGPKELDMVCSPGGTTIEAVRV	242						
QY	248	LEQGGRLRAATMSAVEAATCRAKELSR	273						
DB	243	LEERGFRAAVTEAMTKCKMEKSEALS	268						

RESULT 15

S76367
probable pyrroline-5-carboxylate reductase (EC 1.5.1.2) - *Synechocystis* sp. (strain PCC C;Species: *Synechocystis* sp.

A; Variety: PCC 6803

C:\array\100-000
C:\Date: 25-Apr-1997 #sequence revision 25-Apr-1997 #text change 20-Jun-2000

C:Accession: S76767

Ogawa, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Rikaneke, T.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, O. K.; Okumura, S.; Shimpo, S. *DNA Res.* 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* S.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A: Accession: S76767

A: Accession: S70707
A: Status: nucleic acid sequence not shown: translation not shown

A: Molecule type: DNA

A:Residues: 1-267 <KAN>

A; Cross-references: EMBL:D90916; GB:AB001339; NID:gl653715; PIDN:BAA18679.1; PID:gl65376
A; Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

A; NOC: CIE
C: Genetics:

C; Genetics:
A: Start codon: GTG

A; start codon: GTG
C: superfamily: pyrroline-5-carboxylate reductase

C;superfamily: pyrroline-3-carboxylate reductase

Query Match	29.8%;	Score 412.5;	DB 2;	Length 267;
Best Local Similarity	36.2%;	Pred. No. 5e-25;		
Matches 98: Conservative	50;	Mismatches 110;	Indels 13;	Gaps 4;

QY 10 RVGFV GACRMAGAI AQGLIRAGKVEAQHILASAPTDRLNLCHEQ-ALGCRTHSNOEVLOS 68

4 QIGIIGGWMAEALRLARLAEKTVAPEEIITVGEHPHGARRDYLOKTVYRVSPDNOEAAV 63

QY 69 CLLVIFATKPHVLPVLAEVA-----PVVTEHILVSAAGVSLSTEELLPPNTRVLRV 123

54	SEVLLAVKPOVLDVRLASLAGGNRELV-----TSILAGVSLQRIQGFDDHA-IIRA	116
Db		
124	LPNFCVVQEGAIYVNRGRHHVGSSETKLLQHLLEACRCREVPAYVDIHTGLSSGVAF	183
QY		
117	MENTPATVGAMTATAANKWVEPDLAKAKAIFSAVGNVVEPENLMDAVTGVSGSGPAY	176
Db		
184	VCAPSEALAEAGAKMGKPESSLAHRIAQTLTGATKMLLHEGHPAQLRSDVCTPGGTTTY	243
QY		
177	VALMTEALADGVVLGIPRAAQKLAQTLVLTGAELIKETEERHPAQIKDKVTSPPGTTTA	236
Db		
244	GLHLEQGLRAATWSAVEAATCATKELSRK	274
QY		
237	GVAVLEKMGFRSAITEIAVRAAYRRSQBELGKK	267
Db		

Search completed: July 21, 2003, 09:55:58

Job time : 41 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:53:21 ; Search time 22 Seconds
(without alignments)
1479.099 Million cell updates/sec

Title: US-09-806-536a-14

Perfect score: 1386

Sequence: 1 MAAEPPSRPVGVGAGRMA.....AATSAVEAATCRAKELSRK 274

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA.*

- 1: /cgn2.6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 2: /cgn2.6/ptodata/1/pubpaa/PT_NEW_PUB.pep.*
- 3: /cgn2.6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2.6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2.6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
- 6: /cgn2.6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2.6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2.6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2.6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2.6/ptodata/1/pubpaa/US09_NEW_PUB.pep.1*
- 11: /cgn2.6/ptodata/1/pubpaa/US09_NEW_PUB.pep.2*
- 12: /cgn2.6/ptodata/1/pubpaa/US09_NEW_PUB.pep.3*
- 13: /cgn2.6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 14: /cgn2.6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 15: /cgn2.6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2.6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 17: /cgn2.6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1386	100.0	274	15	US-10-161-418A-11
2	1386	100.0	274	15	US-10-161-418A-13
3	584.5	42.2	319	15	US-10-161-418A-10
4	584.5	42.2	319	15	US-10-161-418A-12
5	560.5	40.4	315	10	US-09-912-717-3
6	537.5	38.8	314	10	US-09-912-717-1
7	477	34.4	319	15	US-10-161-418A-14
8	467.5	33.7	255	11	US-09-925-300-1218
9	408.5	29.5	269	15	US-10-156-761-12258
10	401	28.9	284	15	US-10-128-714-3252
11	401	28.9	284	15	US-10-128-714-8252
12	366	26.2	295	11	US-09-712-363-174
13	362.5	26.2	270	11	US-09-738-626-3960
14	116	8.4	144	10	US-09-939-980-445
15	114	8.2	299	10	US-09-912-020-256
16	95.5	6.9	549	15	US-10-156-761-14029

17	95	6.9	807	12	US-09-930-020A-2	Sequence 2, Appli
18	93.5	6.7	289	15	US-10-156-761-9583	Sequence 9583, Ap
19	93	6.7	2436	10	US-09-795-693-8	Sequence 8, Appli
20	93	6.7	2436	15	US-10-156-239-8	Sequence 8, Appli
21	93	6.7	2436	15	US-10-199-485-8	Sequence 8, Appli
22	92	6.6	408	11	US-09-712-363-273	Sequence 273, App
23	91.5	6.6	452	15	US-10-156-761-12493	Sequence 12493, A
24	89.5	6.5	323	15	US-10-156-761-9372	Sequence 9372, Ap
25	89.5	6.5	336	15	US-10-156-761-10217	Sequence 10217, A
26	89.5	6.5	488	15	US-10-156-761-7949	Sequence 7949, Ap
27	89.5	6.5	858	10	US-09-815-242-11396	Sequence 11396, A
28	89	6.4	447	15	US-10-156-761-9880	Sequence 9880, Ap
29	87.5	6.3	1024	15	US-10-211-962-44	Sequence 44, Appli
30	87.5	6.3	6145	15	US-10-156-761-7962	Sequence 7962, Ap
31	87	6.3	426	15	US-10-156-761-8581	Sequence 8581, Ap
32	87	6.3	949	10	US-09-841-835-10	Sequence 10, Appli
33	87	6.3	1327	10	US-09-841-835-2	Sequence 2, Appli
34	87	6.3	1327	12	US-09-972-115A-8	Sequence 8, Appli
35	86.5	6.2	340	10	US-09-815-242-10852	Sequence 10852, A
36	86.5	6.2	342	10	US-09-815-242-4924	Sequence 4924, Ap
37	86.5	6.2	1835	15	US-10-156-761-7963	Sequence 7963, Ap
38	86.5	6.2	4809	15	US-10-156-761-9090	Sequence 9090, Ap
39	86	6.2	492	10	US-09-815-242-11079	Sequence 11079, A
40	86	6.2	499	15	US-10-156-761-9271	Sequence 9271, Ap
41	85.5	6.2	342	15	US-10-156-761-14291	Sequence 14291, A
42	85.5	6.2	748	10	US-09-815-242-12792	Sequence 12792, A
43	85.5	6.2	792	10	US-09-815-242-12327	Sequence 12327, A
44	85.5	6.2	1687	15	US-10-094-679-3	Sequence 3, Appli
45	85	6.1	1074	10	US-09-509-196A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-161-418A-11
; Sequence 11, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161,418A
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-418A-11

Query Match	100.0%	Score	1386;	DB	15;	Length	274;
Best Local Similarity	100.0%	Pred. No.	1.5e-132;				
Matches	274;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MAAEPPSRPVGVGAGRMAIAOGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTH	60				
Db	1	MAAEPPSRPVGVGAGRMAIAOGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTH	60				
QY	61	SNOEVLQSCLLVIFATKPHVLPVLAIEVAPVVTTHILSVAAAGVSLSTLELLPPNTRV	120				
Db	61	SNOEVLQSCLLVIFATKPHVLPVLAIEVAPVVTTHILSVAAAGVSLSTLELLPPNTRV	120				
QY	121	LRVLNPLPCVVOEGAIVMARGHRVGSSETKLQHLLEACGRCEEVEPYVDIHTGLSGG	180				
Db	121	LRVLNPLPCVVOEGAIVMARGHRVGSSETKLQHLLEACGRCEEVEPYVDIHTGLSGG	180				
QY	181	VAFVCAFEALAEAGVKMGPSSLAHRIAQTLLGTAKMLLHEGHPAQLRSDVCTPGT	240				

```
Db 181 VAFVCAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
QY 241 TIYGLHALEQGLRAATMSAVEAATCRAKELSRK 274
Db 241 TIYGLHALEQGLRAATMSAVEAATCRAKELSRK 274

RESULT 2
US-10-161-418A-13
; Sequence 13, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161,418A
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-418A-13
```

```
Query Match 100.0%; Score 1386; DB 15; Length 274;
Best Local Similarity 100.0%; Pred. No. 1.5e-132;
Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAEPPSRPVGVGAGRMAGTAAGLIRAGKVEAQHILASAPTDNRNLCHFOALGCRTH 60
Db 1 MAAEPPSRPVGVGAGRMAGTAAGLIRAGKVEAQHILASAPTDNRNLCHFOALGCRTH 60

QY 61 SNOEVLQSCLLVIFATKPHVLPVLAELAVPVVTTTEHILVSVAAAGVSLSTLEELPPNTRV 120
Db 61 SNOEVLQSCLLVIFATKPHVLPVLAELAVPVVTTTEHILVSVAAAGVSLSTLEELPPNTRV 120

QY 121 LRVLPNLPCVQVEGAIVMARGRHVGSSETKLLQHLLEACGRCEEPVPEAYVDIHTGLSGS 180
Db 121 LRVLPNLPCVQVEGAIVMARGRHVGSSETKLLQHLLEACGRCEEPVPEAYVDIHTGLSGS 180

QY 181 VAFVCAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
Db 181 VAFVCAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGHPAQLRSDVCTPGGT 240

QY 241 TIYGLHALEQGLRAATMSAVEAATCRAKELSRK 274
Db 241 TIYGLHALEQGLRAATMSAVEAATCRAKELSRK 274
```

```
RESULT 3
US-10-161-418A-10
; Sequence 10, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161,418A
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 319
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-418A-10

Query Match 42.2%; Score 584.5; DB 15; Length 319;
Best Local Similarity 46.4%; Pred. No. 4.5e-51;
Matches 124; Conservative 47; Mismatches 89; Indels 7; Gaps 3;

QY 11 VGFVAGRMAGTAAGLIRAGKVEAQHILASAPTDNRNLCHFOALGCRTHSNOEVLQ 67
Db 3 VGFVAGRMAGTAAGLIRAGKVEAQHILASAPTDNRNLCHFOALGCRTHSNOEVLQ 67

QY 68 SCLLVIFATKPHVLPVLAELAVPVVTTTEHILVSVAAAGVSLSTLEELP---PNTVRVRL 124
Db 62 HSDVLFVAVKPHIIPFILDIEGADIEDRHIVVSCAAGVTISSIEKKLSAFRPAVRVIRCM 121

QY 125 PNLPCVQVEGAIVMARGRHVGSSETKLLQHLLEACGRCEEPVPEAYVDIHTGLSGSVAFV 184
Db 122 TNPVVRVREGATVYATGTHAQVEDGRLEMQLLSTVGCTEVEDLIDAVTGLSGSGPAYA 181

QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGHPAQLRSDVCTPGGTIY 244
Db 182 FTALDALADGGVKGMLPRRLAVRLGAQALLGAAKMLLHSEQHPQLKDNVSSPGGATIH 241

QY 245 LHALEQGLRAATMSAVEAATCRAKEL 271
Db 242 LHVLESGGFRSLLINAVEASCINTREL 268
```

```
RESULT 4
US-10-161-418A-12
; Sequence 12, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161,418A
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-418A-12
```

```
Query Match 42.2%; Score 584.5; DB 15; Length 319;
Best Local Similarity 46.4%; Pred. No. 4.5e-51;
Matches 124; Conservative 47; Mismatches 89; Indels 7; Gaps 3;

QY 11 VGFVAGRMAGTAAGLIRAGKVEAQHILASAPTDNRNLCHFOALGCRTHSNOEVLQ 67
Db 3 VGFVAGRMAGTAAGLIRAGKVEAQHILASAPTDNRNLCHFOALGCRTHSNOEVLQ 67

QY 68 SCLLVIFATKPHVLPVLAELAVPVVTTTEHILVSVAAAGVSLSTLEELP---PNTVRVRL 124
Db 62 HSDVLFVAVKPHIIPFILDIEGADIEDRHIVVSCAAGVTISSIEKKLSAFRPAVRVIRCM 121

QY 125 PNLPCVQVEGAIVMARGRHVGSSETKLLQHLLEACGRCEEPVPEAYVDIHTGLSGSVAFV 184
Db 122 TNPVVRVREGATVYATGTHAQVEDGRLEMQLLSTVGCTEVEDLIDAVTGLSGSGPAYA 181

QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGHPAQLRSDVCTPGGTIY 244
Db 182 FTALDALADGGVKGMLPRRLAVRLGAQALLGAAKMLLHSEQHPQLKDNVSSPGGATIH 241

QY 245 LHALEQGLRAATMSAVEAATCRAKEL 271
Db 242 LHVLESGGFRSLLINAVEASCINTREL 268
```

Db 242 LHVLESGGFRSLINAVEASCIRTEL 268

RESULT 5

US-09-912-717-3

; Sequence 3, Application US/09912717

; Patent No. US20020081691A1

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

Corley, Neil C.

Baughn, Mariah R.

TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE

HOMOLOG

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/912,717

FILING DATE: 24-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/565,910

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Cerrone, Michael C

REGISTRATION NUMBER: 39,132

REFERENCE/DOCKET NUMBER: PF-0532 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-855-0572

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 315 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GENBANK

CLONE: 189498

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-912-717-3

Query Match 40.4%; Score 560.5; DB 10; Length 315;

Best Local Similarity 46.1%; Pred. No. 1.2e-48;

Matches 123; Conservative 46; Mismatches 87; Indels 11; Gaps 5;

QY 11 VGFVAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNCHFOAL---GCRTHSNQEVQLQ 67

Db 3 VGFAGQL--AFAGFTAAGVLAHAKIMASSP-DMDLATYSALKRMGVKLTLPNKEIVQ 59

QY 68 SCLLVIFATKPHVLPVLAIEVAPVVTTEHILVSVAGVSLSTLELLP---PNTRVLRVL 124

Db 60 HSDVLFVAVKPHIIPFLDEIGADIEDRHIVVSCAAGVTISSIEKKLSAFRAPVIRCM 119

QY 125 PNLPCVQBGATVMARGHVGSSSETKLLQHLLEACGCEEVPEAYVDIHTGLSGGVAFV 184

Db 120 TNPVVVREGAVYATGTHAQVEDGRLEQLLSTVGFCTEVEDLIDAVTGLSGGPAY- 178

QY 185 CAFSEALAEAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGTTIYG 244

Db 179 -AFTALDADGGVKGMLPRELAVRLCAQLLGNAAKMLLHSEQHPGQLKNVSSPGGATIIA 237

QY 245 LHALEGGGLRAATMSAVEAANTCRAKEL 271

Db 238 LHVLESGGFRSLINAVEASCIRTEL 264

RESULT 6

US-09-912-717-1

; Sequence 1, Application US/09912717

; Patent No. US20020081691A1

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

Corley, Neil C.

Baughn, Mariah R.

TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE

HOMOLOG

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/912,717

FILING DATE: 24-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/565,910

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Cerrone, Michael C

REGISTRATION NUMBER: 39,132

REFERENCE/DOCKET NUMBER: PF-0532 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-855-0572

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 314 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: PROSNON01

CLONE: 2278458

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-912-717-1

Query Match 38.8%; Score 537.5; DB 10; Length 314;

Best Local Similarity 45.3%; Pred. No. 2.6e-46;

Matches 121; Conservative 45; Mismatches 88; Indels 13; Gaps 5;

QY 11 VGFVAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNCHFOAL---GCRTHSNQEVQLQ 67

Db 3 VGFAGQL--AFAGFTAAGVLAHAKIMASSP-DMDLATYSALKRMGVKLTLPNKEIVQ 59

QY 68 SCLLVIFATKPHVLPVLAIEVAPVVTTEHILVSVAGVSLSTLELLP---PNTRVLRVL 124

Db 58 HSDVLFVAVKPHIIPFLDEIGADIEDRHIVVSCAAGVTISSIEKKLSAFRAPVIRCM 117

QY 125 PNLPCVQBGATVMARGHVGSSSETKLLQHLLEACGCEEVPEAYVDIHTGLSGGVAFV 184

Db 118 TNPVVVREGAVYATGTHAQVEDGRLEQLLSTVGFCTEVEDLIDAVTGLSGGPAY- 176

QY 185 CAFSEALAEAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGTTIYG 244

Db 174 PAYFFYLVAMTDAGILLGLPRDKAHLIVQSAIGAAIMLRDSEGHVPKLRNVTSPAGT 233
QY 241 TIYGLHALEOGGLRAATWSAVEAATCRAKELS 272
Db 234 TINAIRELENHGVRAALIALEAARDRSRELA 265

RESULT 10

US-10-128-714-3252
; Sequence 3252, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3252
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3252

Query Match 28.9%; Score 401; DB 15; Length 284;
Best Local Similarity 37.3%; Pred. No. 1.6e-32;
Matches 101; Conservative 45; Mismatches 115; Indels 10; Gaps 4;

QY 10 RVGFVAGRMAGIAOGLIRAGKVEAQHILASAPTRNLCHFOALGCRTHSNQEVQSC 69
Db 8 KLAFIGGNNMASAIIGLV-SQDMNPANITVSEPDVNRKIAKLGVTQTTNSGEAANA 66
QY 70 LLVIFATKPHVLPVLAELAVPV---VTTEHILVSAAGVSLSTLELLPPN---TRVLR 122
Db 67 DIVIIAVKPTTKNVCOELATANSORTSLPVVVSIAAGITLNSMKEWLRTNDGRTAHIVR 126
QY 123 VLPNLCVVOEGAIVMARGHVGSSSETKLQHLLEACGRCEE--VPEAYVDIHTGLSGSG 180
Db 127 VMPTPALVKEGASGLASDDVTAEKELIGALLOSVSKEATEWVEKEELDDVVTGLSGSG 186
QY 181 VAFVCAFESEALAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
Db 187 PAYFFAMVEHLVASATLGLSKEQATRLAAQTCLGAGKMLVESSEEPQALRKNTVSPNGT 246
QY 241 TIYGLHALEOGGLRAATWSAVEAATCRAKEL 271
Db 247 THAALQTTFESLNFKEIVDKAVQAATSRRAEL 277

RESULT 11

US-10-128-714-8252
; Sequence 8252, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo

; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8252
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8252

Query Match 28.9%; Score 401; DB 15; Length 284;
Best Local Similarity 37.3%; Pred. No. 1.6e-32;
Matches 101; Conservative 45; Mismatches 115; Indels 10; Gaps 4;

QY 10 RVGFVAGRMAGIAOGLIRAGKVEAQHILASAPTRNLCHFOALGCRTHSNQEVQSC 69
Db 8 KLAFIGGNNMASAIIGLV-SQDMNPANITVSEPDVNRKIAKLGVTQTTNSGEAANA 66
QY 70 LLVIFATKPHVLPVLAELAVPV---VTTEHILVSAAGVSLSTLELLPPN---TRVLR 122
Db 67 DIVIIAVKPTTKNVCOELATANSORTSLPVVVSIAAGITLNSMKEWLRTNDGRTAHIVR 126
QY 123 VLPNLCVVOEGAIVMARGHVGSSSETKLQHLLEACGRCEE--VPEAYVDIHTGLSGSG 180
Db 127 VMPTPALVKEGASGLASDDVTAEKELIGALLOSVSKEATEWVEKEELDDVVTGLSGSG 186
QY 181 VAFVCAFESEALAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
Db 187 PAYFFAMVEHLVASATLGLSKEQATRLAAQTCLGAGKMLVESSEEPQALRKNTVSPNGT 246
QY 241 TIYGLHALEOGGLRAATWSAVEAATCRAKEL 271
Db 247 THAALQTTFESLNFKEIVDKAVQAATSRRAEL 277

RESULT 12

US-09-712-363-174
; Sequence 174, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,

```
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 174
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-174

Query Match      26.4%; Score 366; DB 11; Length 295;
Best Local Similarity 31.6%; Pred. No. 6.1e-29;
Matches 90; Conservative 55; Mismatches 116; Indels 24; Gaps 3;

QY 10 RVGFVAGRMAGAGIAGLIRAGKVEAQHILASAPTRDNLCHFOALGCRTHSNQEVLOS 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7 RIAIIGGSGIGALVGLRAGQVLDVLAERMPDRANYLAOTYSVLVT-SAADAVENA 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 LLVIFATKPHVLPVLAIEVAPV-----TTEHILVSVAAAGVSLSTLELLPPNTRVLRV 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 66 TFVVAVKPADVEPTVADLANATAAENDSAQVFTVVAGITIAVFEKLPAGTPVRA 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 LPNLPVQVEGALVARGHVSSTETKLLQHLLEACGRCEVEPYAVVDIHTGLSGGVAF 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 126 MPNAAALVAGVATLAKRFVTPQOLEVSALFDVAVGGVLTVPESQLDAVAVSGGPAY 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 VCAFSEALAEAGVAKMGMPSSLAHRTAAOTLLGTAKMLLHEGQH----- 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 186 FFLVLEALVDAGVGLSRQVADTLAQTMAAGSAMLLERMEDQGGANGELMGLRDLVT 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 227 PAQLRSDVCTPGGTTTLYGLHALEOGGLRAATMSAVEAATCRAKEL 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 246 ASRLRAAVTSPGGTTAAALRELERGGFRGFMVADAAVQAQAKSRSEQL 290
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-09-738-626-3960
; Sequence 3960, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3960
; LENGTH: 270

; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3960

Query Match      26.2%; Score 362.5; DB 11; Length 270;
Best Local Similarity 30.6%; Pred. No. 1.2e-28;
Matches 83; Conservative 63; Mismatches 114; Indels 11; Gaps 4;

QY 11 VGFVAGRMAGAGIAGLIRAGKVEAQHILASAPTRDNLCHFOALGCRTHSNQEV 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4 IAVIGGGIGALVGLI-AANMNPQIRV---TNRSEERGQELDRYGLNMTDNSQAA 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 QSCLLVIFATKPHVLPVLAIEVAPV---TTEHILVSVAAAGVSLSTLELLPPNTRVLRV 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 60 DEADVFLCVKPKFIVEVLSEITGTLDNNSAQSVVSMAGISIAAMEESASAGLPVVRV 119
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 LPNLPVQVEGALVARGHVSSTETKLLQHLLEACGRCEVEPYAVVDIHTGLSGGVAF 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 120 MPNTPMLVKGOMSTVTKGRYVDAEQLEQVKDILLSTVGDVLEVAESDIDAVTAMSGSPAY 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 VCAFSEALAEAGVAKMGMPSSLAHRTAAOTLLGTAKMLLHEGQHQAQLRSDVCTPGGTTIY 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 180 LFLVTEALIEAGVNLGLPRATAKLAVASPEGAATMMKETGKPPSELRAAGVSSPAGTTVA 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 GLHLEOGGLRAATMSAVEAATCRAKELSRK 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 240 AIRELESINGAFYRAAQACADRSEELGKR 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-09-939-980-445
; Sequence 445, Application US/09939980
; Patent No. US20020082234A1
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; Burnham, Martin
; Hodgson, John
; Knowles, David
; Lonetto, Michael
; Nicholas, Richard
; Pratt, Julie
; Reichard, Richard
; Rosenberg, Martin
; Ward, Judith
; TITLE OF INVENTION: No. US20020082234A1el Prokaryotic Polynucleotides,
; Polypeptides and Their Uses
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/939,980
; FILING DATE: 27-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/936,165
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
```



```

?       TELEX: <unknown>
?
? INFORMATION FOR SEQ ID NO: 445:
?
?     SEQUENCE CHARACTERISTICS:
?
?       LENGTH: 144 amino acids
?       TYPE: amino acid
?       STRANDEDNESS: single
?       TOPOLOGY: linear
?
?       MOLECULE TYPE: Protein
?
?       SEQUENCE DESCRIPTION: SEQ ID NO: 445:
US-09-939-980-445

```

Query Match 8.4%; Score 116; DB 10; Length 144;
Best Local Similarity 26.8%; Pred. No. 0.00054;
Matches 30; Conservative 24; Mismatches 56;
Indels 2; Gaps 2;

Qy	16	AGRMAGAIAGLIRACKVEAQHI-LASAPTRDNI-CHF-QALGCRTHSHSQEVLQSCLLVI	73
Db	12	AGNMAAIFTGTTINSNLDANDIYITNKSNEQAKAFAEKLGVNYSYDDATLLKQADYVF	71
Qy	74	FATRPVHLPAVLAEAPVVTTEHILVSVAAAGVSLTLEELPNTRVRLVP	125
Db	72	LCTRPFDALATRIKPHITKDXCFMSIMAGIPTDYIXOOLECONXPARTMP	123

```

RESULT 15
US-09-912-020-256
; Sequence 256, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.00IDV1
; CURRENT APPLICATION NUMBER: US/09/912.020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492.709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 256
; LENGTH: 299

```

; LIFE. PRI
 ; ORGANISM: E. Coli
 US-09-912-020-256

Query Match 8.2%; Score 114; DB 10; Length 299;
Best Local Similarity 21.7%; Pred. No. 0.0024;
Matches 69; Conservative 46; Mismatches 119; Indels 84; Gaps 16;

[illegible]

```

Qy 214 L-GTAKMLLHEGQHFA-----QLRSDVCTPGG-----TTIYGLHALEQGLRA 255
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 220 LDKAKPMVDMRFKPGFTRDLHIKLANALDTSHGVGAQLPLTAAYVMMQALRADGLGT 279
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 256 ATMSAVEAATCRAKELSR 273
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 280 ADHSAL---ACYYEKLAK 294
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Search completed: July 21, 2003, 09:57:02
Job time : 24 secs

```

Search completed: July 21, 2003, 09:57:02
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:42:26 ; Search time 71 Seconds
(without alignments)
514.235 Million cells

Title: US-09-806-536A-14

Perfect score: 1386
Sequence: 1 MAAAEPSPRRVGFVAGRMA.....AATMSAVEAATCRAKELSRK 274

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

Database :

A_Geneseq_101002:*

1: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1980.DAT:**

2: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1981.DAT:**

3: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1982.DAT:**

4: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1983.DAT:**

5: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1984.DAT:**

6: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1985.DAT:**

7: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1986.DAT:**

8: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1987.DAT:**

9: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1988.DAT:**

10: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1989.DAT:**

11: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1990.DAT:**

12: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1991.DAT:**

13: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1992.DAT:**

14: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1993.DAT:**

15: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1994.DAT:**

16: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1995.DAT:**

17: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1996.DAT:**

18: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1997.DAT:**

19: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1998.DAT:**

20: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA1999.DAT:**

21: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA2000.DAT:**

22: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA2001.DAT:**

23: /SIDS2/cgcdata/geneseq/geneseqp-emb1/AA2002.DAT:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1386	100.0	274	21	AAY92517	Human OXRE-14	Ho
2	1386	100.0	274	22	AAB95591	Human protein sequ	
3	1383	99.8	274	22	AAG66956	Human dihydropyrr	
4	590.5	42.6	343	23	ABP41278	Human ovarian anti	
5	584.5	42.2	319	22	ABP47480	Human p5CR protein	
6	577.5	41.7	320	22	AAE132784	Human delta 1-pyrr	
7	560.5	40.4	273	22	ABE61855	Drosophila melanog	
8	560.5	40.4	315	22	AAE12785	Human delta 1-pyrr	
9	560.5	40.4	320	22	AAB47779	Human Py-CR protei	
10	537.5	38.8	314	21	AAB20584	Human delta 1-pyrr	

ALIGNMENTS

RESULT 1

RECEIVED
AAY92517
ID AAY92517 standard; Protein; 274 AA.

AA
AC AAY92517:

10-AUG-2000 (first entry)

XX
DE
Human OXRE-14.XX
KW OXPF-14: oxidoreductase: pyrroline-5-carboxylate-reductase:

antiproliferative; anticancer; hepatotropic; antiviral; antiasthmatic;
nootropic; neuroprotectant; antiparkinsonian's; antisclerotic;
axiolytic; antischizophrenic; anti-inflammatory; antiarthritic.

XX
05
Homo sapiens.

XX	Key	Location/Qualifiers
FH		

Region

FT Modified-site

Modified-site

FT Modified-site

Modified-site
FT
FT

ET
ET
ET
Modified-site

Modified-site

100

Human delta 1-pyr
Drosophila melanog
Arabidopsis thalia
Listeria monocytog
Human prostate can
Arabidopsis thalia
Escherichia coli p
Propionibacterium
Human protein sequ
Mycobacterium tube
C glutamicum prote
Corynebacterium gl
Lactococcus lactis
Streptococcus poly
Novel human enzyme
Novel human enzyme
Streptococcus poly
Arabidopsis thalia
S. epidermidis ope
Staphylococcus epi
N meningitidis pr
Novel human diagn
Listeria monocytog
Human reductase-ll
Human secreted pro
Arabidopsis thalia
Novel human diagn
pyrroline-5-carbox.
E. coli proliferat
Human ATP-binding
Human MPR-related
Human ATP-binding
Human ATP-binding
Human ATP-binding
Human ATP-binding

FT Modified-site 259
 FT /note= "potential phosphorylation site"
 FT Modified-site 265
 FT /note= "potential phosphorylation site"
 XX
 XX
 PN WO200020604-A2.
 XX 13-APR-2000.
 XX
 PF 06-OCT-1999; 99WO-US23434.
 XX
 PR 06-OCT-1998; 98US-0172227.
 PR 02-DEC-1998; 98US-0155202.
 PR 10-MAR-1999; 99US-0123911.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Lal P, Guegler KJ, Gorgone GA, Corley NC, Baughn MR, Tang YT;
 PI Hillman JL, Bandman O, Azimzai Y, Au-Young J, Yue H, Lu DAM;
 PI Yang J;
 XX
 DR WPI: 2000-303785/26.
 DR N-PSDB: AAA09388.
 XX
 PT Purified polypeptide for treating or preventing disorders associated
 PT with decreased expression or activity of oxidoreductase molecules
 XX
 PS Claim 1; Page 84; 97pp; English.
 XX
 CC This OXRE-14 has identity with pyrroline-5-carboxylate-reductase.
 CC The polypeptides are useful for treating or preventing a disorder
 CC associated with decreased expression or activity of OXRE. Antagonists of
 CC OXRE are useful for treating or preventing a disorder associated with
 CC increased expression or activity of OXRE. The disorders include cell
 CC proliferative disorders (cirrhosis, hepatitis), cancer (leukemia,
 CC melanoma), hypopituitarism and hyperpituitarism, hypothyroidism and
 CC hyperthyroidism, metabolic disorders (Addison's disease, cystic
 CC fibrosis), reproductive disorders (infertility, ovulatory defects),
 CC neurological disorders (Alzheimer's disease, Parkinson's disease,
 CC multiple sclerosis), mental disorders (anxiety, schizophrenia),
 CC autoimmune/inflammatory disorders (acquired immunodeficiency syndrome
 CC (AIDS), asthma, osteoarthritis), and viral infections. The
 CC polynucleotides may be used in Southern or Northern analysis, polymerase
 CC chain reaction (PCR), or in enzyme-linked immunosorbent assays (ELISA).
 XX
 SQ Sequence 274 AA;
 Query Match 100.0%; Score 1386; DB 21; Length 274;
 Best Local Similarity 100.0%; Pred. No. 5.7e-131;
 Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAAEPSPRVGFVAGRMAGATAOGLIRAGKVEAQHILASAPTDRLNLCFHQALGCRTH 60
 DB 1 MAAAEPSPRVGFVAGRMAGATAOGLIRAGKVEAQHILASAPTDRLNLCFHQALGCRTH 60
 QY 61 SNOEVLQSCLLVIFATKPHVLPVLAEAPVVTTEHILVSAAGVSLSTLELLPPNTRY 120
 DB 61 SNOEVLQSCLLVIFATKPHVLPVLAEAPVVTTEHILVSAAGVSLSTLELLPPNTRY 120
 QY 121 LRVLPNLCVQVQEGALVMARGHVSSETKLLQHLLEACGRCVEEPEAYVDIHTGLSGG 180
 DB 121 LRVLPNLCVQVQEGALVMARGHVSSETKLLQHLLEACGRCVEEPEAYVDIHTGLSGG 180
 QY 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAATLLGTAKMLLHGHPAQLRSDVCTPGGT 240
 DB 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAATLLGTAKMLLHGHPAQLRSDVCTPGGT 240
 QY 241 TYIYGLHALBQGLRAATMSAVEAATCRAKELSRK 274
 DB 241 TYIYGLHALBQGLRAATMSAVEAATCRAKELSRK 274

RESULT 2

AAB95591
 ID AAB95591 standard; Protein; 274 AA.
 XX
 AC AAB95591;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ.ID NO:18269.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 99JP-0248036.
 PR
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs.
 XX
 PS Claim 8; SEQ ID 18269; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 274 AA;
 Query Match 100.0%; Score 1386; DB 22; Length 274;
 Best Local Similarity 100.0%; Pred. No. 5.7e-131;
 Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAAEPSPRVGFVAGRMAGATAOGLIRAGKVEAQHILASAPTDRLNLCFHQALGCRTH 60
 DB 1 MAAAEPSPRVGFVAGRMAGATAOGLIRAGKVEAQHILASAPTDRLNLCFHQALGCRTH 60
 QY 61 SNOEVLQSCLLVIFATKPHVLPVLAEAPVVTTEHILVSAAGVSLSTLELLPPNTRY 120

Db 61 SNOEVLOSCLLVIFATKPHVLPVLAIEVAPVWVTEHILVSVAGMSLSTLEELLPPNTRV 120
QY 121 LRVLPNLPVVOEGAIIVMARGHRVGSSETKLLQHLLEACGRCCEVPEAYVDIHTGLSGSG 180
Db 121 LRVLPNLPVVOEGAIIVMARGHRVGSSETKLLQHLLEACGRCCEVPEAYVDIHTGLSGSG 180
QY 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
Db 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
QY 241 TIYGLHALEOGGLRAATMSAVEAATCRAKELSRK 274
Db 241 TIYGLHALEOGGLRAATMSAVEAATCRAKELSRK 274

RESULT 3

AAG66956
ID AAG66956 standard; Protein; 274 AA.

XX
AC AAG66956;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human dihydropyrrrole-5-carboxylate reductase 30 polypeptide.
XX
KW Human; dihydropyrrrole-5-carboxylate reductase 30; cancer; cytostatic;
KW human immunodeficiency virus; HIV; infection; immunological disease;
KW inflammatory disease.
XX
OS Homo sapiens.
XX
PN CN1298002-A.
XX
PD 06-JUN-2001.
XX
PF 24-NOV-1999; 99CN-0124090.
XX
PR 24-NOV-1999; 99CN-0124090.
XX
PA (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.
XX
PI Mao Y, Xie Y;
XX
WPI: 2001-489680/54.
DR N-PSDB; AAH77597.

PT Human dihydropyrrrole-5-carboxylate reductase 30 as one new kind of
PT polypeptide and polynucleotides encoding this polypeptide -
XX
PS Claim 1: Page 20-21 (disclosure); 26pp; Chinese.
XX
CC The invention relates to a novel polypeptide, human
CC dihydropyrrrole-5-carboxylate reductase 30, polynucleotides encoding
CC this polypeptide and a DNA recombination process to produce the
CC polypeptide. The polypeptide is useful for treating various diseases,
CC such as malignant tumours, nosohaemia, HIV infection, immunological
CC diseases and inflammatory diseases. The invention also provides an
CC antibody against the polypeptide. The present sequence is the
CC polypeptide of the invention.

XX
SQ Sequence 274 AA;

Query Match 99.88; Score 1383; DB 22; Length 274;
Best Local Similarity 99.68; Pred. No. 1.1e-130;
Matches 273; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAREPSRRVGFVAGRMAGATAQGLIRAGKVEAOHILASAPTDRLNLCHEQALGCRTH 60
Db 1 MAAREPSRRVGFVAGRMAGATAQGLIRAGKVEAOHILASAPTDRLNLCHEQALGCRTH 60
QY 61 SNOEVLOSCLLVIFATKPHVLPVLAIEVAPVWVTEHILVSVAGMSLSTLEELLPPNTRV 120
|||||

Db 61 SNOEVLOSCLLVIFATKPHVLPVLAIEVAPVWVTEHILVSVAGMSLSTLEELLPPNTRV 120
QY 121 LRVLPNLPVVOEGAIIVMARGHRVGSSETKLLQHLLEACGRCCEVPEAYVDIHTGLSGSG 180
Db 121 LRVLPNLPVVOEGAIIVMARGHRVGSSETKLLQHLLEACGRCCEVPEAYVDIHTGLSGSG 180
QY 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
Db 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
QY 241 TIYGLHALEOGGLRAATMSAVEAATCRAKELSRK 274
Db 241 TIYGLHALEOGGLRAATMSAVEAATCRAKELSRK 274

RESULT 4

ABP41278
ID ABP41278 standard; Protein; 343 AA.

XX
AC ABP41278;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HOGDC67, SEQ ID NO:2410.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory system disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antinflammatory; gynaecological; reproductive; chromosome 17.

XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PR 07-JUN-2000; 2000US-209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
WPI: 2002-147878/19.
DR N-PSDB; ABQ54355.

PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 11: SEQ ID No 2410; 2922pp; English.

XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic

sequence shown in sequence listing of the specification"
188
/note= "This amino acid residue is absent in the
sequence shown in sequence listing of the specification"
189
/note= "This amino acid residue is absent in the
sequence shown in sequence listing of the specification"
190
US6268192-B1.
31-JUL-2001.
05-MAY-2000; 2000US-0565910.
18-JUN-1998; 98US-0099676.
(INCY-) INCYTE GENOMICS INC.
Hillman JL, Corley NC, Baughn MR;
WPI; 2001-647056/74.
N-PSDB; AAD20894.
New delta 1-pyrroline-5-carboxylate reductase polypeptides and
polynucleotides, useful for diagnosing, treating and preventing
neural disorders, connective tissue disorders or disorders of cell
proliferation -
Claim 1; Fig 2; 32pp; English.
The present invention relates to delta 1-pyrroline-5-carboxylate
reductase (P5CR) polypeptides and polynucleotides. The P5CR gene is used
in gene therapy. The P5CR nucleic acid and amino acid sequences are
useful in the diagnosis, treatment and prevention of neuronal disorders
(e.g. akathisia, Alzheimer's disease, amnesia, or amyotrophic lateral
sclerosis), connective tissue disorders (e.g. cystic fibrosis, myocardial
fibrosis, or osteoporosis) and disorders of cell proliferation (e.g.
actinic keratosis, arteriosclerosis bursitis, cancers). P5CR may also be
used for screening libraries of compounds in various drug screening
techniques. The present sequence is human delta 1-pyrroline-5-carboxylate
reductase homologue (P5CRH).

Query Match 41.7%; Score 577.5; DB 22; Length 320;
Best Local Similarity 46.4%; Pred. No. 1.6e-49;
Matches 124; Conservative 48; Mismatches 88; Indels 7; Gaps 3;
QY 11 VGVFVAGRMAGATAGLIRAGKVEAQHILASAPTDRNLCHFQAL---GCRTHSNQEVLYQ 67
DB 3 VGFAGQALAYALARGFTAGILSAHKIIASSP-EMNLPTVSALRKMGNLTRSNKETVK 61
QY 68 SCILVIFATKPHVLPVLAEPVVTTEHILSVSAAGVSLSTLELL---PPNTRVLRLV 124
DB 62 HSDVFLAVKPHIIPILDEIGADVQARHIVVSCAAGVTISSEKKLMAFAPAPKVRM 121
QY 125 PNLPCVQGAIVMARGHRVGSSETKLQHLLEACGRCEEPVAVVDIHTGLSGGVAFV 184
DB 122 TNPVVOGATVIATGTHALVEDGQLLEQLMSSVGFCTEVEDLIDAVTGLSGGPAYA 181
QY 185 CAFSEALAGVAKMGMPSSLAHRIAQTLLGTAKMLLHEGOHPAQLRSDVCTPGGTTIYG 244
DB 182 FMALDALADGGVKMGLPRRLAIQLGAQALLGAAKMLLDSEHQPCQLKDNVSPGATITHA 241
QY 245 LHAEOGGLRAATMSAVEAATCRANEL 271
DB 242 LHFESGGFRSLINAVEASCINTREL 268

RESULT 7
ABB61855
ID ABB61855 standard; Protein: 273 AA.
XX

AC ABB61855;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 12357.
DE Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS WO200171042-A2.
XX 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL05958.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -
XX Disclosure; SEQ ID NO 12357; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 273 AA;
SQ Query Match 40.4%; Score 560.5; DB 22; Length 273;
Best Local Similarity 44.2%; Pred. No. 6.6e-48;
Matches 118; Conservative 50; Mismatches 96; Indels 3; Gaps 3;
QY 10 RVGFVAGRMAGATAGLIRAGKVEAQHILASA-PTDR-NLCHFQALGCRTHSNQEVLYQ 67
DB 6 KTGFLGGGNKAKALAGFLAAGLAKPNTLIASVHPADKLSLQSFQSLGVETVKNAPVQ 65
QY 68 SCILVIFATKPHVLPVLAEPVVTTEHILSVSAAGVSLSTLELLPPNTRVLRLV 127
DB 66 QSDVFEVSKPVQVPSVLSLSEIQP-LSSGKFLSVAMGTLTSTIESLSPQARVIRMPNL 124
QY 128 PCVVOGATVIMARGHRVGSSETKLQHLLEACGRCEEPVAVVDIHTGLSGGVAFV 187
DB 125 PAVVCGSVFVVGSKATDADITQKLLQSVGTCEPVEDSOLDVVTALSGSPAYVFM 184
QY 188 SEALAGVAKMGMPSSLAHRIAQTLLGTAKMLLHEGOHPAQLRSDVCTPGGTTIYG 247
DB 185 IEALADGAVHMGMPRLAYRLASQTVLGAGHVMRDSGMPGQKDGVTSPAGSTAAALRQ 244
QY 248 LEOGGLRAATMSAVEAATCRANELSRK 274
DB 245 LELSGFRAAVSGAVEQATLCRQISCK 271

RESULT 8

AAE12785
ID AAE12785 standard; Protein; 315 AA.
AC
XX
XX AAE12785;
DT
XX
XX 15-JAN-2002 (first entry)
DE Human delta 1-pyrroline-5-carboxylate reductase (P5CR).
XX
XX Human; delta 1-pyrroline-5-carboxylate reductase; P5CR; cystic fibrosis;
KW osteoporosis; neuronal disorder; gene therapy; akathisia; drug screening;
KW actinic keratosis; Alzheimer's disease; amyotrophic lateral sclerosis;
KW connective tissue disorder; myocardial fibrosis; cell proliferation;
KW arteriosclerosis bursitis; cancer; amnesia; neuroprotective; cytostatic;
KW antiarteriosclerotic; osteopathic; cardiac.
XX
XX Homo sapiens.
XX
XX US6268192-B1.
PN
XX
XX 31-JUL-2001.
PD
XX
XX 05-MAY-2000; 2000US-0565910.
PF
XX
XX 18-JUN-1998; 98US-0099676.
PR
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX
XX Hillman JL, Corley NC, Baughn MR;
FI
XX
XX WPI; 2001-647056/74.
DR
XX
XX New delta 1-pyrroline-5-carboxylate reductase polypeptides and
PT polynucleotides, useful for diagnosing, treating and preventing
PT neuronal disorders, connective tissue disorders or disorders of cell
PT proliferation
XX
XX Disclosure; Fig 2; 32pp; English.
XX
XX The present invention relates to delta 1-pyrroline-5-carboxylate
CC reductase (P5CR) polypeptides and polynucleotides. The P5CR gene is used
CC in gene therapy. The P5CR nucleic acid and amino acid sequences are
CC useful in the diagnosis, treatment and prevention of neuronal disorders
CC (e.g. akathisia, Alzheimer's disease, amnesia, or amyotrophic lateral
CC sclerosis), connective tissue disorders (e.g. cystic fibrosis, myocardial
CC fibrosis, or osteoporosis) and disorders of cell proliferation (e.g.
CC actinic keratosis, arteriosclerosis bursitis, cancers). P5CR may also be
CC used for screening libraries of compounds in various drug screening
CC techniques. The present sequence is human delta 1-pyrroline-5-carboxylate
CC reductase (P5CR).
XX
XX
SQ Sequence 315 AA;
Query Match 40.4%; Score 560.5; DB 22; Length 315;
Best Local Similarity 46.1%; Pred. No. 8e-48;
Matches 123; Conservative 46; Mismatches 87; Indels 11; Gaps 5;
QY 11 VGFVAGRMAGAIAGLIRAGKVEAQHILASAPTRNLCHFOAL---GCRTHSNQEVILQ 67
Db 3 VGFIGAGQL--AFAGKGTAAAGVLAHAKIMASSP-DMDLATVSALRKMGVLTLPNKETVQ 59
QY 68 SCLLVIFATKPHVLPVLAIEAPVVTTEHILVSVAGVSLSTLEELLP---PNTNRLVRL 124
Db 60 HSDVFLAVKPHIIPFILDIGADIEDRHVVSCAAGVTISSIEKLSAFRAPVIRCM 119
QY 125 PNLPCVQGEAIVMARGHVGSSETKLLOHLLACGCEVEPPEAYVDIHGLSSGVAFV 184
Db 120 TNPVVRGATVYATGTHAQVEDGLMEQLLSTLVGFTCEVEEDLDIVATGLSGSPAY- 178
QY 185 CAFSEALAEAGVKMGMPSSLAHRIAATLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYG 244
Db 179 -AFTALDADGGVKMGLPRRLAVRLGAQALGAAKMLLHSEHPQLKDNVSSPGGATIHA 237

QY 245 LHALEQGGRLRAATMSAVEAATCRAKEL 271
Db 238 LHVLESGGFRSLLINAVEASCIRTREL 264
RESULT 9
AAB74779
ID AAB74779 standard; Protein; 320 AA.
XX
XX AAB74779;
DT 06-JUN-2001 (first entry)
DE Human Py-CR protein SEQ ID NO:4.
XX
XX Human; pyrroline 5'-carboxylate reductase; Py-CR; P5CR.
XX
XX Homo sapiens.
OS
XX
XX CN1274728-A.
PN
XX
XX 29-NOV-2000.
PD
XX
XX 25-MAY-1999; 99CN-0107071.
PF
XX
XX 25-MAY-1999; 99CN-0107071.
PR
XX
XX (UYFU-) UNIV FUDAN.
PA
XX
XX Yu L, Fu O, Zhang H;
PI
XX
XX WPI; 2001-211749/22.
DR
XX
XX N-PSDB; AAF81847.
DR
XX
XX New human protein and its code sequence, preparation and application -
PT
XX
XX Claim 4; Page 16; 20pp; Chinese.
PS
XX
XX The present invention describes a human protein designated Py-CR, which
CC is homologous to the human pyrroline-5'-carboxylate reductase (P5CR)
CC (EC1.5.1.2). The present invention also describes methods for the
CC application and production process of the py-CR polynucleotide and
CC protein sequences. The present sequence represents the human Py-CR
CC protein as given in the present invention.
XX
XX
SQ Sequence 320 AA;
Query Match 40.4%; Score 560.5; DB 22; Length 320;
Best Local Similarity 45.7%; Pred. No. 8.2e-48;
Matches 122; Conservative 48; Mismatches 90; Indels 7; Gaps 3;
QY 11 VGFVAGRMAGAIAGLIRAGKVEAQHILASAPTRNLCHFOAL---GCRTHSNQEVILQ 67
Db 3 VGFIGAGQLANALARGFTAAAGILSAHKIIASSP-EMNLPTVSALRKMGVNLTRSKETVK 61
QY 68 SCLLVIFATKPHVLPVLAIEAPVVTTEHILVSVAGVSLSTLEELLP---PNTNRLVRL 124
Db 62 HSDVFLAVKPHIIPFILDIGADYQARHIVVSCAAGVTISSVEKKLMAFQAPKVIKCM 121
QY 125 PNLPCVQGEAIVMARGHVGSSETKLLOHLLACGCEVEPPEAYVDIHGLSSGVAFV 184
Db 122 TNPVVRGATVYATGTHALVEGQLEQLLMSSVGFTCEVEEDLDIVATGLSGSPAYA 181
QY 185 CAFSEALAEAGVKMGMPSSLAHRIAATLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYG 244
Db 182 FVALDALADGGVKMGLPRRLAIQGAQALGAAKMLLHSEHPQLKDNVSSPGGATIHA 241
QY 245 LHALEQGGRLRAATMSAVEAATCRAKEL 271
Db 242 LHVLESGGFRSLLINAVEASCIRTREL 268
RESULT 10

us-09-806-536a-14.rag

Mon Jul 28 08:07:02 2003

PR	09-MAR-1999;	99US-0123548.	PR	15-JUL-1999;	99US-0144005.
PR	23-MAR-1999;	99US-0125788.	PR	16-JUL-1999;	99US-0144085.
PR	29-MAR-1999;	99US-0126264.	PR	16-JUL-1999;	99US-0144086.
PR	29-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144331.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144332.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	99US-0129845.	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	99US-0130077.	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	99US-0130449.	PR	19-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.	PR	22-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134219.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134221.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145918.
PR	19-MAY-1999;	99US-0134941.	PR	27-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	99US-01351124.	PR	28-JUL-1999;	99US-0145951.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146386.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146388.
PR	25-MAY-1999;	99US-0136021.	PR	02-AUG-1999;	99US-0146389.
PR	27-MAY-1999;	99US-0136392.	PR	03-AUG-1999;	99US-0147038.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147204.
PR	01-JUN-1999;	99US-0137222.	PR	04-AUG-1999;	99US-0147302.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147192.
PR	07-JUN-1999;	99US-0137502.	PR	05-AUG-1999;	99US-0147260.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147303.
PR	08-JUN-1999;	99US-0138094.	PR	06-AUG-1999;	99US-0147416.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138847.	PR	09-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	99US-0139113.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139452.	PR	11-AUG-1999;	99US-0148319.
PR	17-JUN-1999;	99US-0139453.	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139456.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149226.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151065.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151066.
PR	22-JUN-1999;	99US-0139819.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151338.
PR	23-JUN-1999;	99US-0140354.	PR	07-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142577.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.

```

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 28-OCT-1999; 99US-0162143.

Query Match 34.3%; Score 475; DB 21; Length 276;
Best Local Similarity 40.0%; Pred. No. 2.7e-39;
Matches 106; Conservative 54; Mismatches 103; Indels 2; Gaps 2;

QY 10 RVGFVAGRMAGAIAGLIIRAGKVEAQHILASAPTDRLMCH-FOALGCRTHHSNQEVLS 68
DB 12 KVGFAGAGMAESIARGVAVSGVLPNRCITAVHSNLRNRRDVFSGVNFSTSEEVYKE 71
QY 69 CLIVIPATKPHVLPVLAEPVVTTHILVSVAGVSLSTLELLPNTNRLVRLNLP 128
DB 72 SDVVFESKPVQVKKAVTELKSKNLILVSVAGIKLNDLQE-WGQDREIRVWNT 130
QY 129 CVVQEGAIVMARGHVSSETKLLOHLEACGRCCEVEPEAYVDIHTGLSGSGVAFVCAFS 188
DB 131 AAVGEAASVMSLGTGATEEDGAIVAMLFAGVKGILKADKRMFDVATVGLSGSGPAYIFLAI 190
QY 189 EALAEAGVAKMGPPSSLAHRIAOTLLGTAKMLLHEGHPAOLRSDVCTPGGTTIYGLHAL 248
DB 191 EALADGVAAVGLPRELALSLASQTVLGAATWVSKTGKHPGVKDDVTSPPGTTIAGVHEL 250
QY 249 EGGGLRAATMSAVEAATCRAKELSR 273
DB 251 EKGFSRATLMNAVVAAKRSRELSQ 275

RESULT 14
ABB47427
ID ABB47427 standard; Protein; 266 AA.
XX
AC ABB47427;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #131.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.
XX

```

```

PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-PR01118.
XX
PR 11-APR-2000; 2000FR-0004629.
XX
PA (INSP ) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Solana JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss R;
XX
DR WPI; 2002-010914/01.
XX
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides
XX
PS Claim 6; SEQ ID No 132; 192pp; French.
XX
CC The present invention relates to the genome sequence of Listeria
CC monocytogenes EGB-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms,
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccine compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 266 AA;

Query Match 33.8%; Score 469; DB 23; Length 266;
Best Local Similarity 38.8%; Pred. No. 1e-38;
Matches 101; Conservative 55; Mismatches 96; Indels 8; Gaps 2;

QY 10 RVGFVAGRMAGAIAGLIIRAGKVEAQHILASAPTDRLMCHFOAL-----GCRTHHSNOE 64
DB 3 RIGFTGAGMGTAMIRGLAOLANLVKREIIVGG---RNLEKLPLEAEFTGLQITDTEK 59
QY 65 VLOSCLLVIFATKPHVLPVLAEPVVTTHILVSVAGVSLSTLELLPNTNRLVRL 124
DB 60 LVEQADIIILAVKPYTPEILTSVKEKLPDKIIISVAGVTTQDLBELTSATKIVRM 119
QY 125 PNLPCVQEGAIVMARGHVSSETKLLOHLEACGRCCEVEPEAYVDIHTGLSGSGVAFV 184
DB 120 PNTPALVGEAMSSVSPNTNVTSEELKEVTAIFTSGEAEVVSLENLDAVIGVSGSPAYV 179
QY 185 CAFSEALAEAGVAKMGPPSSLAHRIAOTLLGTAKMLLHEGHPAOLRSDVCTPGGTTIY 244
DB 180 YMFIALADGAVLSGMPDRKAYKFAAQAVLGAATVLEGTGEHFGKLDKMTSPGGTTIEA 239
QY 245 LHALEQGGGLRAATMSAVEAA 264
DB 240 VKSLENDGFRSAVINAVQAA 259

```

QY 256 ATMSAVEATCRAKEL 271
Db 189 LLINAVEASCIRTREL 204

Search completed: July 21, 2003, 09:53:14
Job time : 73 secs

RESULT 15
AAB56640
ID AAB56640 standard; Protein: 255 AA.
XX
AC AAB56640;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen protein sequence SPQ ID NO:1218.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 200WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
DR N-PSDB; AAF15843.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 11; Page 1636-1637; 2338pp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 255 AA;
Query Match 33.7%; Score 467.5; DB 21; Length 255;
Best Local Similarity 48.5%; Pred. No. 1.4e-38;
Matches 95; Conservative 34; Mismatches 64; Indels 3; Gaps 1;
QY 79 HVLPAVLAEPVPTTEHILVSVAGVSLSTLEELLP---PNTRYRLVLPNLPVVOEGA 135
Db 9 HIIPFLDEIGADIEDRHIVVSCAGVTISSEKKLSAFRPAFRVIRMTNTPVVVREGA 68
QY 136 IVMARGHVGSSTKLLQHLLEACGCEEPYAYVDIHTGLSGGVAFVCAFEALAEGA 195
Db 69 TVATGTHAQVEGRLEMEQLLSVGFCTEVEDLIDAVTGLSGSGPAYAFTALDALDGG 128
QY 196 VKMGMPSSLAHRTAOTLLGTAKMLLHEGHPAQLRSVDVCTPGGTTIYGLHLEOGGLRA 255
Db 129 VKMGLPRRLAVRLGAQALLGAARKMLLSHQHGPGLKDNVSPGGATIHVHLES GGFRS 188

II	AAH14370 standard; cDNA; 2615 BP.	Db	2555	GGGGGGCCATCGCGCAGGGGCTCATCAGACAGAGAAAGTGAAGCTCAGCACATAC	2496
XX					
AC	AAH14370;	QY	147	GCCAGTGACCAACACAGACAGAGAACCTATGTCTACCTTTCAAGCTCTGGGTTGCCGACACG	206
XX					
DT	26-JUN-2001 (first entry)	Db	2495	GCCAGTGACCAACACAGACAGAGAACCTATGTCTACCTTTCAAGCTCTGGGTTGCCGACACG	2436
XX					
DE	Human cDNA sequence SEQ ID NO:11778.	QY	207	CACCTCAACACAGGAGGTGCTGCAGAGTGCCTGCTGCTGCTATCTTTGCCAACAGCCTCAT	266
XX					
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.	Db	2435	CACCTCAACACAGGAGGTGCTGCAGAGTGCCTGCTGCTGCTATCTTTGCCAACAGCCTCAT	2376
XX					
OS	Homo sapiens.	QY	267	GTGCTGCCAGCTGTCTTGGCAGAGTGGCTCTCTGTGTTCCACCTGACACATCTTGGTG	326
XX					
PN	EP1074617-A2.	Db	2375	GTGCTGCCAGCTGTCTTGGCAGAGTGGCTCTCTGTGTTCCACCTGACACATCTTGGTG	2316
XX					
PD	07-FEB-2001.	QY	327	TCCGTGGCTGCTGGGGTGTCTCTGAGCACCTGAGGAGTCTGCTGCCCCCAACACACG	386
XX					
PF	28-JUL-2000; 2000EP-0116126.	Db	2315	TCCGTGGCTGCTGGGGTGTCTCTGAGCACCTGAGGAGTCTGCTGCCCCCAACACACG	2256
XX					
PR	29-JUL-1999; 99JP-0248036.	QY	387	GTGCTGCGGGTCTTGGCCCAACCTGCTGTGTGGTCCAGGAAGGGGCCATAGTATGGCG	446
XX					
PR	27-AUG-1999; 99JP-0300253.	Db	2255	GTGCTGCGGGTCTTGGCCCAACCTGCTGTGTGGTCCAGGAAGGGGCCATAGTATGGCG	2196
XX					
PR	11-JAN-2000; 2000JP-0118776.	QY	447	CGGGGCGGCCAGTGGGGAGCAGCAGACCAAGCTCTGACACATCTGCTGGAGGCTCTG	506
XX					
PR	02-MAY-2000; 2000JP-0183767.	Db	2195	CGGGGCGGCCAGTGGGGAGCAGCAGACCAAGCTCTGACACATCTGCTGGAGGCTCTG	2136
XX					
XX	09-JUN-2000; 2000JP-0241899.				
PA	(HELI-) HELIX RES INST.				
XX					
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	QY	507	GGCGGTGTGAGGAGTGTCTGAAGCTTACGTGACATCCACACTGGCCTCAGTGGCACT	566
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	Db	2135	GGCGGTGTGAGGAGTGTCTGAAGCTTACGTGACATCCACACTGGCCTCAGTGGCACT	2076
XX					
DR	WPI; 2001-318749/34.	QY	567	GGCGTGGCCTTCTGTGTGCTCCGAGGCGCTTGGCTGAAGAGCCGTCAAGATGGCG	626
XX					
PT	Primer sets for synthesizing polynucleotides, particularly the 5602	Db	2075	GGCGTGGCCTTCTGTGTGCTCCGAGGCGCTTGGCTGAAGAGCCGTCAAGATGGCG	2016
PT	full-length cDNAs defined in the specification, and for the detection	QY	627	ATGCCAGCAGCCTGCGCCACCGATCGCTGCCAGACCTGCTGGGGAGCGGCAAGATG	686
PT	and/or diagnosis of the abnormality of the proteins encoded by the	Db	2015	ATGCCAGCAGCCTGCGCCACCGATCGCTGCCAGACCTGCTGGGGAGCGGCAAGATG	1956
XX					
PS	Claim 8; SEQ ID 11778; 2537pp + CD ROM; English.	QY	687	CTGCTGACAGGCGCCAAACACCCAGCCAGCTCGCTCAGAGCTGTCACCCCGGGTGGC	746
XX					
CC	The present invention describes primer sets for synthesizing 5602	Db	1955	CTGCTGACAGGCGCCAAACACCCAGCCAGCTCGCTCAGAGCTGTCACCCCGGGTGGC	1896
CC	full-length cDNAs defined in the specification. Where a primer set	QY	747	ACCACATCTATGACTCCAGCCCTGGAGCGGGGCTGGAGAGCCGTCAAGATGGCG	806
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary	Db	1895	ACCACATCTATGACTCCAGCCCTGGAGCGGGGCTGGAGAGCCGTCAAGATGGCG	1836
CC	to the complementary strand of a polynucleotide which comprises one of	QY	807	GCGTGGAGCTGCCACTGCGGGCCAAAGAGCTCAGCAGAAAGTAGGCTGGGCTTGG	866
CC	the 5602 nucleotide sequences defined in the specification, where the	Db	1835	GCGTGGAGCTGCCACTGCGGGCCAAAGAGCTCAGCAGAAAGTAGGCTGGGCTTGG	1776
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination	QY	867	CCATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	926
CC	of an oligonucleotide comprising a sequence complementary to the	Db	1775	CCATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1716
CC	complementary strand of a polynucleotide which comprises a 5'-end	QY	927	GCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT	986
CC	sequence and an oligonucleotide comprising a sequence complementary to a	Db	1715	GCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT	1656
CC	polynucleotide which comprises a 3'-end sequence, where the	QY	987	CAGGGGCGAGGACTTGGGAGGTTCACAGCAGCGGGGAGCCCGACAGTGGGGGACACTC	1046
CC	oligonucleotide comprises at least 15 nucleotides and the combination of	Db	1655	CAGGGGCGAGGACTTGGGAGGTTCACAGCAGCGGGGAGCCCGACAGTGGGGGACACTC	1596
CC	the 5'-end sequence/3'-end sequence is selected from those defined in	QY	1047	CTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT	1106
CC	the specification. The primer sets can be used in antisense therapy and	Db	1595	CTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT	1536
CC	in gene therapy. The primers are useful for synthesizing polynucleotides,	QY	1107	CCACTTGTCTCAACATTTGGTCTGAGGGGCCCAAGAGATGGCTCTTGGTTCATTTGCC	1166
CC	particularly full-length cDNAs. The primers are also useful for the	Db	1535	CCACTTGTCTCAACATTTGGTCTGAGGGGCCCAAGAGATGGCTCTTGGTTCATTTGCC	1476
CC	detection and/or diagnosis of the abnormality of the proteins encoded by	QY	1167	GCATGTTGGGCACTTGGTGTGAGGCCATGAACAGAACTTACGGTAACAGGACCGGCTGGC	1226
CC	the full-length cDNAs. The primers allow obtaining of the full-length	Db	1475	GCATGTTGGGCACTTGGTGTGAGGCCATGAACAGAACTTACGGTAACAGGACCGGCTGGC	1416
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and				
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to				
CC	AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632				
CC	represent oligonucleotides, all of which are used in the exemplification				
CC	of the present invention.				
XX					
SQ	Sequence 2615 BP; 492 A; 831 C; 818 G; 474 T; 0 other;				
	Query Match 96.5%; Score 1291.8; DB 22; Length 2615;				
	Best Local Similarity 99.1%; Pred. No. 1.9e-278;				
	Matches 1299; Conservative 0; Mismatches 12; Indels 0; Gaps 0;				
QY	27 AAGATGGCAGCTCGCGAGCGCTCTCGCGCGCGCGTGGCTTCTGTTGGCGCGCGCGCATG 86				
Db	2615 AAGATGGCAGCTCGCGAGCGCTCTCGCGCGCGCGTGGCTTCTGTTGGCGCGCGCGCATG 2556				
QY	87 GCGGGGCGCATCGCGAGGGCTTCATCAGACGAGGAAAGTGAAGCTCAGCACATCTG 146				

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
DR P-PSDB; AAM78866.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX Claim 1; Page 1961-1966; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.

XX Sequence 5055 BP; 989 A; 1736 C; 1512 G; 818 T; 0 other;

Query Match 54.0%; Score 723; DB 22; Length 5055;
Best Local Similarity 95.9%; Pred. No. 1.3e-151;
Matches 750; Conservative 0; Mismatches 30; Indels 2; Gaps 1;

QY 45 CCGTCTCCGCGGGGTGGCTTCGTGGCGGGCGCGCATGGCGGGGCCATCGCGCAG 104
DB 15 CCGTCTCCGCGGGGTGGCTTCGTGGCGGGCGCGCATGGCGGGGCCATCGCGCAG 72
QY 105 GGCTCATCAGACGAGGAAAGTGAAGCTCAGCACATAGTGGCGAGTGCACCAACAGAC 164
DB 73 GGCTCATCAGACGAGGAAAGTGAAGCTCAGCACATAGTGGCGAGTGCACCAACAGAC 132
QY 165 AGAACCTATGCTATTTCAAGCTCTGGGTTCCGGGACACCGCATCTCAACAGGAGGTG 224
DB 133 AGGAACCTATGCTATTTCAAGCTCTGGGTTCCGGGACACCGCATCTCAACAGGAGGTG 192
QY 225 CTGCAGAGTGGCTGCTGCTCATCTTTGCCACCAAGCCCTCATGCTGCCAGCTCTCCTG 284
DB 193 CTGCAGAGTGGCTGCTGCTCATCTTTGCCACCAAGCCCTCATGCTGCCAGCTGCTCCTG 252

QY 285 GCAGAGTGGCTCTGTGGTTCACACTGAACACATCTTGGTCCGTGGCTGCTGGGGTG 344
DB 253 GCAGAGTGGCTCTGTGGTTCACACTGAACACATCTTGGTCCGTGGCTGCTGGGGTG 312
QY 345 TCTGTAGCACCCCTGGAGGAGCTGCTGCCCCCAACACACACGGGTGCTGGGGTCTTGGCC 404
DB 313 TCTGTAGCACCCCTGGAGGAGCTGCTGCCCCCAACACACACGGGTGCTGGGGTCTTGGCC 372
QY 405 AACCTGCCCTGTGTGGTCCAGGAGGGCCATAGTATGGCGGGGCCCGCCAGCTGGGG 464
DB 373 AACCTGCCCTGTGTGGTCCAGGAGGGCCATAGTATGGCGGGGCCCGCCAGCTGGGG 432
QY 465 AGCAGCAGACCAAGCTCTGTCAGCATCTGTCGAGGCCCTGTGGCGGTGTGAGGAGGTG 524
DB 433 AGCAGCAGACCAAGCTCTGTCAGCATCTGTCGAGGCCCTGTGGCGGTGTGAGGAGGTG 492
QY 525 CCTGAAGCCTACGTCGACATCCACATGGGCTCAGTGGCAGTGGCGGTGCTGGTGTGT 584
DB 493 CCTGAAGCCTACGTCGACATCCACATGGGCTCAGTGGCAGTGGCGGTGCTGGTGTGT 552
QY 585 GCATTCTCCGAGGCCCTGGCTGAGGAGCGCTCAAGATGGCATGCCACGACGCTGGCC 644
DB 553 GCATTCTCCGAGGCCCTGGCTGAGGAGCGCTCAAGATGGCATGCCACGACGCTGGCC 612
QY 645 CACCGCATCGTGGCCAGACCCCTGCTGGGACGCCCAAGATGCTGTCACAGGGGCCAA 704
DB 613 CACCGCATCGTGGCCAGACCCCTGCTGGGACGCCCAAGATGCTGTCACAGGGGCCAA 672
QY 705 CACCCAGCCAGCTGGCGCTCAGACGTGTGACCCCGGTGGCACCACCATCTATGGACTC 764
DB 673 CACCCAGCCAGCTGGCGCTCAGACGTGTGACCCCGGTGGCACCACCATCTATGGACTC 732
QY 765 CACCCCTGGAGCAGCGGGGCTGCGAGCAGCCACCATAGCGCGGTGGAGGCTGCCACC 824
DB 733 CACCCCTGGAGCAGCGGGGCTGATGCCAAGCCTGGGAGGCTGGGAGCTGGGGTC 792
QY 825 TG 826
DB 793 TG 794
RESULT 7
AAS42169
ID AAS42169 standard; DNA; 5647 BP.
XX AAS42169;
AC AAS42169;
XX 17-DEC-2001 (first entry)
XX Genomic sequence #485 encoding novel human enzyme polypeptide.
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ds.
OS Homo sapiens.
XX WO200155301-A2.
PN 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01239.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465566/50.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases
XX
PS Disclosure; SEQ ID No 2295; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the

functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. Influenza). The polynucleotides of the invention can also be used in gene therapy. AAS41685-AAS42192 represent DNA sequences encoding for the novel human enzyme polypeptides of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 5647 BP; 1011 A; 1640 C; 1768 G; 1226 T; 2 other;

Query Match 44.8%; Score 599.4; DB 22; Length 5647;
Best Local Similarity 96.9%; Pred. No. 4.8e-124;
Matches 653; Conservative 0; Mismatches 17; Indels 4; Gaps 4;

664 CCCTCTGGGACGCGCAAGATGCTGTGACGAGGCGCAACACCCAGCCGAGTGGCT 723
675 CTCCTCAGGGGACGGCGAAGATGCTGTGACGAGGCGCAACACCCAGCCGAGTGGCT 734

724 CAGAGGTGTGACCCCGGGTGGACACCATCTATGACATCCAGCCCTGAGGAGGGG 783
735 CAGAGGTGTGACCCCGGGTGGACACCATCTATGACATCCAGCCCTGAGGAGGGG 794

784 GCCTCGAGGACGACCATGAGCGCGGTGGAGGTGCCACCTGCGGGGCCAAGAGCTCA 843
795 GNCCTCGAGCAG-CACCATGAGCGCGGTGGAGG-TGCCACCTGCGGGGCCAAGAGCTCA 852

844 GCAGAAAGTAGGCTGGGCTCTGGGCATCCTTTCTGCTGCTCTGTGCGCCCTGCTCCCTG 903
853 GCAGAAAGTAGGCTGGGCTCTGG-CATCCTTTCTGCTGCTGTGCGCCCTGCTCCCTG 911

904 TGCTCCCTTCCCTGAGGACTGGGCTCCCTCCCTCTCATGAGGCTCTCTACTGTCC 963
912 TGCTCCCTTCCCTGAGGACTGGGCTCCCTCCCTCTCATGAGGCTCTCTACTGTCC 971

964 TTCTCCCTTGCACAGGAAATGACAGGGGCGAGGACTTGGAGGTTCCAGAGCGGGGG 1023
972 TTCTCCCTTGCACAGGAAATGACAGGGGCGAGGACTTGGAGGTTCCAGAGCGGGGG 4031

1024 AGCCCGGACAGTGGGGACACTCTCTCCCTCCAGTGAGGAGGACCGTGGTGGTGG 1083
4032 AG-CCGACAGTGGGGACACTCTCTCCCTCCAGTGAGGAGGACCGTGGTGGTGG 4090

1084 CTCGCCCTTCTCAGTGAGGACCATCTGCTGCAACATTTGTTCTCAGGGGCCCAAGA 1143
4091 CTCGCCCTTCTCAGTGAGGACCATCTGCTGCAACATTTGTTCTCAGGGGCCCAAGA 4150

1144 GATGGCGTCTTGCTCATTTGCCCGCATGTTGGCGAGTGGTTGAGGCCATGAACAGAAC 1203
4151 GATGGCGTCTTGCTCATTTGCCCGCATGTTGGCGAGTGGTTGAGGCCATGAACAGAAC 4210

1204 TTACGGGTACAGGCGGCTGCCCAATGCCCTGGTCTGGAGCTGGAGCTTGCTTTGGCT 1263
4211 TTACGGGTACAGGCGGCTGCCCAATGCCCTGGTCTGGAGCTTGAGCTTTGGCT 4270

1264 TTCCAGTGGGCTGTGACAGTACAGCCAGCGCGGCTGCTCATCTCAGCTCTAGGGGGC 1323
4271 TTCCAGTGGGCTGTGACAGTACAGCCAGCGCGGCTGCTCATCTCAGCTCTAGGGGGC 4330

1324 ACAGCATATGGGG 1337
4331 ACAGGCATATGGG 4344

AAH06264
ID AAH06264 standard; cDNA; 563 BP.
XX
AC AAH06264;
DT
XX 26-JUN-2001 (first entry)
XX Human cDNA clone (5'-primer) SEQ ID NO:3099.
DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
OS
XX EP1074617-A2.
PN
XX 07-FEB-2001.
PD
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX 29-JUL-1999; 99JP-0248036.
PR
XX 27-AUG-1999; 99JP-0300253.
PR
XX 11-JAN-2000; 2000JP-0118776.
PR
XX 02-MAY-2000; 2000JP-0183767.
PR
XX 09-JUN-2000; 2000JP-0241899.
PR
XX (HELI-) HELIX RES INST.
PA
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
DR
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 3099; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 563 BP; 99 A; 171 C; 181 G; 108 T; 4 other;

Query Match 36.4%; Score 486.6; DB 22; Length 563;
Best Local Similarity 95.9%; Pred. No. 4e-99;
Matches 540; Conservative 0; Mismatches 18; Indels 5; Gaps 4;

27 AAGATGGCAGCTGC-GGAGCCGCTCCGCGCGCGTTCGTGGCGCGGCGCGCAT 85
1 AAGATGGCAGCTGGCGGAGCGCTCCGCGCGCGTTCGTGGCGCGGCGCGCAT 60

86 GCGCGGGCCCATCGCGCAGGCGCTCATCAGAGCAGGAAAGTGAAGTCAAGCATACT 145


```
Db      61  GCGCGGGGCCATCGCGAGGCGCTCATCAGCAGGAGAAAGTGAAGCTCAGCACATACT 120
Qy      146  GGCAGTGCACACACACAGGAACTATGCTACTTTCAGCTCTGGTTGCCGACAC 205
Db      121  GGCAGTGCACACACAGGAACTATGCTACTTTCAGCTCTGGTTGCCGACAC 180
Qy      206  GCACTCCAAACAGAGTGTCTGAGAGTGCCTGCTGCTCATCTTTGCCACCAAGCTCA 265
Db      181  GCACTCCAAACAGAGTGTCTGAGAGTGCCTGCTGCTCATCTTTGCCACCAAGCTCA 240
Qy      266  TGTGCTGCCAGCTGTCTCTGG- CAGAGTGGCTCCTGTGFTCAACACTGAACATCTTGG 324
Db      241  TGTGCTGCCAGCTGTCTCTGGACAGAGTGGCTCCTGTGFTCAACACTGAACATCTTGG 300
Qy      325  TGTCCGTGCTGTCTGGGTGTCTCTGAGCACCTGTGAGGAGCTGTGCCCCCAACACAC 384
Db      301  TGTCCGTGCTGTCTGGGTGTCTCTGAGCACCTGTGAGGAGCTGTGCCCCCAACACAC 360
Qy      385  GGTGCTGCGGTCTTGCCCAACTGCCCTGTGTGTCTCCAGGAAGGGCCATAG--TGAT 442
Db      361  GGTGCTGCGGTCTTGCCCAACTGCCCTGTGTGTCTCCAGGAAGGGCCATAGATNATG 420
Qy      443  GCGCGGGGCGCCAGTGGGAGCAGCAGACCAAGCTCTTCGACGATCTCTGTGGAGGC 502
Db      421  CCTCGGGGCGCCAGTGGGAGCAGCAGACCAAGCTCTTCGACGATCTCTGTGGATGC 480
Qy      503  CTGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGACATCCACACTGGCTCAGTGG 562
Db      481  CTGTGGCGGTGTGAGGAGTGCCTGAATCCTACGTGACATCCACACTGGCTCAATGG 540
Qy      563  CAGTGG-CGTGGCTTCGTGTGT 584
Db      541  CAGTGGCTGTGCGCTTCGNGTNT 563

RESULT 9
AAS41342
ID  AAS41342 standard; cdNA; 270 BP.
XX
AC  AAS41342;
XX
DT  17-DEC-2001 (first entry)
XX
DE  cDNA encoding novel human enzyme polypeptide #558.
XX
KW  Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW  ligase; hyperproliferative disorder; immunodeficiency disorder;
KW  autoimmune disorder; neurological disorder; metabolic disorder;
KW  inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW  blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW  anti arthritic; nephrotropic; anticoagulant; ss.
XX
OS  Homo sapiens.
XX
PN  WO20015301-A2.
XX
PD  02-AUG-2001.
XX
PF  17-JAN-2001; 2001WO-US01239.
XX
PR  31-JAN-2000; 2000US-0179065.
PR  04-FEB-2000; 2000US-0180628.
PR  24-FEB-2000; 2000US-0184664.
PR  02-MAR-2000; 2000US-0186350.
PR  16-MAR-2000; 2000US-0189874.
PR  17-MAR-2000; 2000US-0190076.
PR  18-APR-2000; 2000US-0198123.
PR  19-MAY-2000; 2000US-0205515.
PR  07-JUN-2000; 2000US-0209467.
PR  28-JUN-2000; 2000US-0214886.
PR  30-JUN-2000; 2000US-0215135.
PR  07-JUL-2000; 2000US-0216647.
PR  07-JUL-2000; 2000US-0216880.
PR  11-JUL-2000; 2000US-0217487.
PR  11-JUL-2000; 2000US-0217496.
PR  14-JUL-2000; 2000US-0218290.
PR  26-JUL-2000; 2000US-0220963.
PR  26-JUL-2000; 2000US-0220964.
PR  14-AUG-2000; 2000US-0224518.
PR  14-AUG-2000; 2000US-0224519.
PR  14-AUG-2000; 2000US-0225213.
PR  14-AUG-2000; 2000US-0225214.
PR  14-AUG-2000; 2000US-0225266.
PR  14-AUG-2000; 2000US-0225267.
PR  14-AUG-2000; 2000US-0225268.
PR  14-AUG-2000; 2000US-0225270.
PR  14-AUG-2000; 2000US-0225447.
PR  14-AUG-2000; 2000US-0225757.
PR  14-AUG-2000; 2000US-0225758.
PR  14-AUG-2000; 2000US-0225759.
PR  18-AUG-2000; 2000US-0226279.
PR  22-AUG-2000; 2000US-0226681.
PR  22-AUG-2000; 2000US-0226888.
PR  22-AUG-2000; 2000US-0227182.
PR  23-AUG-2000; 2000US-0227009.
PR  30-AUG-2000; 2000US-0228924.
PR  01-SEP-2000; 2000US-0229287.
PR  01-SEP-2000; 2000US-0229343.
PR  01-SEP-2000; 2000US-0229344.
PR  01-SEP-2000; 2000US-0229345.
PR  05-SEP-2000; 2000US-0229509.
PR  05-SEP-2000; 2000US-0229513.
PR  06-SEP-2000; 2000US-0230437.
PR  08-SEP-2000; 2000US-0231242.
PR  08-SEP-2000; 2000US-0231243.
PR  08-SEP-2000; 2000US-0231244.
PR  08-SEP-2000; 2000US-0231413.
PR  08-SEP-2000; 2000US-0231414.
PR  08-SEP-2000; 2000US-0232080.
PR  12-SEP-2000; 2000US-0231968.
PR  14-SEP-2000; 2000US-0232397.
PR  14-SEP-2000; 2000US-0232398.
PR  14-SEP-2000; 2000US-0232399.
PR  14-SEP-2000; 2000US-0232400.
PR  14-SEP-2000; 2000US-0232401.
PR  14-SEP-2000; 2000US-0233063.
PR  14-SEP-2000; 2000US-0233064.
PR  21-SEP-2000; 2000US-0233065.
PR  21-SEP-2000; 2000US-0234223.
PR  25-SEP-2000; 2000US-0234274.
PR  25-SEP-2000; 2000US-0234997.
PR  26-SEP-2000; 2000US-0234998.
PR  26-SEP-2000; 2000US-0235484.
PR  27-SEP-2000; 2000US-0235834.
PR  29-SEP-2000; 2000US-0235836.
PR  29-SEP-2000; 2000US-0236327.
PR  29-SEP-2000; 2000US-0236367.
PR  29-SEP-2000; 2000US-0236368.
PR  29-SEP-2000; 2000US-0236369.
PR  29-SEP-2000; 2000US-0236370.
PR  02-OCT-2000; 2000US-0236802.
PR  02-OCT-2000; 2000US-0237037.
PR  02-OCT-2000; 2000US-0237038.
PR  02-OCT-2000; 2000US-0237039.
PR  13-OCT-2000; 2000US-0239935.
PR  13-OCT-2000; 2000US-0239937.
PR  20-OCT-2000; 2000US-0240960.
PR  20-OCT-2000; 2000US-0241221.
PR  20-OCT-2000; 2000US-0241785.
PR  20-OCT-2000; 2000US-0241786.
PR  20-OCT-2000; 2000US-0241787.
PR  20-OCT-2000; 2000US-0241808.
```


PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-465566/50.
XX P-PSDB; AAU23472.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing,
XX preventing, treating neural, immune system, muscular, reproductive,
XX pulmonary, cardiovascular, renal, proliferative disorders and cancerous
XX diseases.
XX
XX Claim 4; SEQ ID NO 568; 1180pp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
XX polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
XX encoding them. The enzyme polypeptides of the invention may comprise the
XX functional classes of oxidoreductases, transferases, hydrolases, lyases,
XX isomerases or ligases. The sequences of the invention are useful in the
XX diagnosis, treatment, prevention and/or prognosis of a wide range of
XX disorders including hyperproliferative disorders (e.g. cancer),

CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAS40785-AAS41684 represent cDNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 270 BP; 54 A; 82 C; 88 G; 45 T; 1 other;
SQ

Query Match 18.1%; Score 241.8; DB 22; Length 270;
Best Local Similarity 95.9%; Pred. No. 1.2e-44;
Matches 258; Conservative 1; Mismatches 8; Indels 2; Gaps 1;

Qy 10 AGTCGTCCGAGGCAACAAGATGCGAGCTCGGAGCGCTCTCCGGCGCGGTGGCTTCG 69
|||
Db 2 AGGCGTCCGAGGCAACAAGATGCGAGCTCGGAGCGCTCTCCGGCGCGGTGGCTTCG 61
|||
Qy 70 TGGCGCGGCGCGCATGGCGGGGCCATCGCGAGGCCCTCATCAGAGAGGAAAGTGG 129
|||
Db 62 TGGCGCGGCGCGCATGGCGGGGCCATCGCGAGGCCCTCATCAGAGAGGAAAGTGG 121
|||
Qy 130 AAGCTCAGCACATACCTGCGCCAGTGCACCAACAGACAGAACCTATGTCACTTTCAAGCTC 189
|||
Db 122 AAGCTCAGCACATACCTGCGCCAGTGCACCAACAGACAGAACCTATGTCACTTTCAAGCTC 181
|||
Qy 190 TGGGTTCGCGGACCAACGACCTCCCAACAGGAGGTGCTGCAGAGCT--GCCTGCTCGTCAT 247
|||
Db 182 TGGGTTCGCGGACCAACGACCTCCCAACAGGAGGTGCTGCAGAGCTGCTGCTGCTCGTCAT 241
|||
Qy 248 CTTTGCCACCAAGCCTCATGTGTGCCAG 276
|||
Db 242 CTTTGCCACCAAGCCTCATGTGTGCCAG 270
|||

RESULT 10
AEN95903
ID AEN95903 standard; DNA: 1792 BP.
XX
XX AC AEN95903;
XX
XX DT 13-AUG-2002 (first entry)
XX
XX DE Gene #2401 used to diagnose liver cancer.
XX
XX KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX KW metastatic liver tumour; cytostatic; expression profile; disease state;
XX KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
XX OS Homo sapiens.
XX
XX PN WO200229103-A2.
XX
XX PD 11-APR-2002.
XX
XX PF 02-OCT-2001; 2001WO-US30589.
XX
XX PR 02-OCT-2000; 2000US-237054P.
XX
XX PA (GENE-) GENE LOGIC INC.
XX
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI: 2002-426119/45.
XX
XX DR Diagnosing and detecting the progression of liver cancer,
XX PT hepatocellular carcinoma or metastatic liver tumor in a patient,
XX PT involves detecting the level of expression of two or more genes in a

PT liver tissue sample -

XX Claim 1; SEQ ID NO 2401; 298pp; English.

XX The invention relates to a novel method for diagnosing and detecting the progression of liver cancer, hepatocellular carcinoma or metastatic liver tumour in a patient, and differentiating metastatic liver cancer from hepatocellular carcinoma in a patient, involving detecting the level of expression of two or more genes represented in ABN93503-ABN97455 in a tissue sample. The method of the invention has hepatotropic, and cytostatic activity. The method is useful for diagnosing and detecting the progression of liver cancer, hepatocellular carcinoma and metastatic liver carcinoma in a patient. The method is useful for identifying expression profiles which serve as useful diagnostic markers as well as markers that can be used to monitor disease states, disease progression, drug toxicity, drug efficacy and drug metabolism.

XX Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1792 BP; 342 A; 573 C; 522 G; 355 T; 0 other;

Query Match 16.8%; Score 224.8; DB 24; Length 1792;
Best Local Similarity 58.0%; Pred. No 1.2e-40;
Matches 466; Conservative 0; Mismatches 317; Indels 21; Gaps 3;

QY 56 GCGGTGGGCTTCGTGGGCGGCGCGCATGGCGGGGCCATCGCGAGGGGCTCATCAG 115
DB 14 GAGCGTGGGCTTCATGGGCGCTGGCGAGCTGCTTTTGGCCTGGCGAAGGGCTTCACAGC 73
QY 116 AGCAGGAAGAAGTGAAGCTCAGACATAGTGCAGTGCACCAACAGACAGAACCTATG 175
DB 74 AGCAGGGCTTGGCTGGCCCAAGAGATAATGCTAGCTCCC---CAGACATGAGCTGGC 130
QY 176 TCACCTTCAAGCTC-----TGGGTTGCGGACACGACACTCCACACGAGGCTGCT 226
DB 131 CACAGTTTCTGCTCTCAGGAAGATGGGCTGAAGTTGACACCCACACAGAGACGGT 190
QY 227 GCAGAGCTGCTGCTGCTATCTTGGCCACCAAGCTCATGTGTCGACAGTGTCTCTGGC 286
DB 191 GCAGCACAGTATGCTCTTCTGCTGTGAAGCCACACATCATCCCTTTCATCTGGA 250
QY 287 AGAGTGGCTTCCTGTGGTCCACCACTGAACACATCTTGGTGTCCGTGCTGCGGTGTC 346
DB 251 TGAATAGGCGCGACATTGAGACAGACACATTTGGTGTCTGCGGCGCGGCGTCAAC 310
QY 347 TCTGACACCTTGGAGGAGCTGCTG-----CCCCCAACACACGCGGTCTGCGGGT 397
DB 311 CATCAGCTCCATTGAGAGAGAGCTGTACAGCGTTTTCGGCCAGCCCGCAGGGTTCATCCGCTG 370
QY 398 CTTGCCCCAACCTGCTGTGTGTCAGGAGGGGCCATAGTATGCGCGGGCGGCCCA 457
DB 371 CATGACCAACATCTCAGTGTGTCGGGAGGGGGCCACCCTGTATGCCACAGGCACGCA 430
QY 458 CGTGGGAGCAGCAGACCAAGCTCTCTGACGATCTGCTGGAGGCTGTGGCGGTGTGA 517
DB 431 CGCCAGGTGGAGGAGCGGAGGCTCATGGAGAGCTGTGAGCAGGTTGGCTTCTGCAC 490
QY 518 GCAGTGTCTGAAGCTTAGCTGCATCCACTGGCTCAGTGGCTAGTGGCGAGTGGCGCTT 577
DB 491 GGAGGTGGAAGAGGACCTGATGTATGTCGCTACGGGGCTCAGTGGCAGCGGCCCGCCTA 550
QY 578 CTTGTGTGCTATCTCCGAGGCGCTGCTGCTGAAGAGCGCTCAAGATGGCATGGCCACGAG 637
DB 551 CGCATTCACAGCCCTGGATGCCCTGCTGATGGGGGTGTGAAGATGGGACTTCAAGGCG 610
QY 638 CTTGCGCCACCGCATCTGCTGCCAGACCTCTGCTGGGACGCCCAAGATGCTGTGCACGA 697
DB 611 CTTGTCAGTTCGCGCTCGGGGCGCCAGGCGCTCTCTGGGGCTGCCAAGATGCTGTGCACT 670
QY 698 GGGCCCAACACCCAGCCAGCTGCCTCAGAGCTGTGACCCCGGGTGGCACCACCATCTA 757
DB 671 AGAACAGCACCAGCCAGCTCAAGGACACGCTCAGCTCTCTCTGTTGGGGCCACCATCCA 730

QY 758 TGGACTCCACGCCCTGGAGCAGGCGGGCTGGGACGACCCACCATGACCGCGTGGAGGC 817
DB 731 TGCCCTGTGATGTGCTGGAGAGTGGGGGCTTCGGCTCCTGCTCATCAACAGCTGTGGAGGC 790
QY 818 TGCCACCTCCCGGGGCCAAGAGGAGCT 841
DB 791 CTCCTGTCATCCGCACACGCGGAGCT 814

RESULT 11

ABQ54355/c

ID ABQ54355 standard; cDNA; 1889 BP.

XX AC ABQ54355;

XX DT 22-AUG-2002 (first entry)

XX DE Human ovarian antigen HOGDC67 cDNA, SEQ ID NO:235.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antinflammatory; gynaecological; reproductive; chromosome 17;
KW gene; ss.

XX OS Homo sapiens.

XX PN WO200200677-A1.

XX PD 03-JAN-2002.

XX PF 07-JUN-2001; 2001WO-US18569.

XX PR 07-JUN-2000; 2000US-209467P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Birse CE, Rosen CA;

XX WPI; 2002-147878/19.

XX P-PSDB: ABP41278.

XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -

XX Claim 1; SEQ ID NO 235; 2922pp; English.

XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,

CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents cDNA encoding a human ovarian antigen of the
CC invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1889 BP; 364 A; 554 C; 589 G; 379 T; 3 other;

Query Match 16.8%; Score 224.8; DB 24; Length 1889;
Best Local Similarity 58.0%; Pred. NO. 1.2e-40;
Matches 466; Conservative 0; Mismatches 317; Indels 21; Gaps 3;

QY 56 GCGCGTGGCTTCGTGGGGGGGGCGGCATGCGGGGGCCATCGCGCAGGCGCTCATCAG 115
DB 1813 GAGCGTGGCTTCATCGCGCGTGGCCAGCTGCTTTTCCCTGGCCAAAGGCTTCACAGC 1754
QY 116 AGCAGGAAAAGTGAAGCTCAGCACATCTAGCCAGTGCACCAACAGACAGAGAACTATG 175
DB 1753 AGCAGGCGTCTTGGCTGCCCAAGATATGGCTAGCTCC---CAGACATGGACCTGGC 1697
QY 176 TCACTTTCAAGTCT-----TGGGTTGCGGGACACAGCACTCCAAACCAGGAGGTGCT 226
DB 1696 CACAGTTTCTGCTCTCAGGAAGATGGGGGTGAAGTTGACAGCCCAACAAGAGACGGT 1637
QY 227 CGAGAGTCCCTGCTCGTCATCTTCCACCAAGCTCATGTGTCAGCTGCTGCTGCTGCGC 286
DB 1636 GCAGCAGCATGATGTGCTCTTCTGGCTGTGAAGCCACACATCATCCCTTCATCCTGGA 1577
QY 287 AGAGTGGCTCCTGTGGTCCACCATGAAACATCTTTGGTGTCCGTGGCTGTGGGGTGTG 346
DB 1576 TGAATAGCGCGGACATTTAGGACAGACACATTTGGTGTCTCGCGGGCGGGGTAC 1517
QY 347 TCTGAGCACCTTGGAGGAGCTGCTG-----CCCCCAACACACAGGCTGCTGCGGGT 397
DB 1516 CATCAGCTCCATTGAGAAGACTGTACAGCGTTTCGGCCAGCCCGCAGGCTCATCCGTG 1457
QY 398 CTTGCCCACTTCCCTGTGTGCTCCAGGAAGGGCCATAGTGTGGCGGGCGGCCA 457
DB 1456 CATGACCAACATCTCAGTCTGTGTCGGGAGGGGGCCACCGTGTATGCCACAGGACGCA 1397
QY 458 CGTGGGGAGCAGCAGACAAAGCTCTCTGCAGCATCTGCTGGAGGCTGTGGCGGGTGTGA 517
DB 1396 CGCCAGGTGGAGGAGCGGAGGCTCATGGAGCAGCTGTGTAGCAGCGTGGGCTTCTGCAC 1337
QY 518 GGAGTGTCTGAAGCTTACCTAGTCATCCACACTGGCTCAGTGGCAGTGGGCTGGCTT 577
DB 1336 GGAGTGTGAAGAGGAGCTGATTTGATGCGCTCAGGGGGTCAGTGCGACGGCGCCGCTA 1277
QY 578 CGTGTGTGATTTCTCCGAGGCGCTGCTGAAGAGCGCTCAAGATGGGCATGCCACGACG 637
DB 1276 CGCATTCACAGCCCTGGATGCCCTGGCTGATGGGGGTGTGAAGTGGGACTTCAAGCGG 1217
QY 638 CTTGGCCCAACCCGATCGCTGCGCAGACCGCTGTGCGGAGCGGCCAAGATGCTGCTGCACGA 697
DB 1216 CTTGGCAGTCCGCTCGGGGGCCAGGCCCTCTCGGGGGTGTGCAAGATGCTGCTCACTC 1157
QY 698 GGGCCCAACCCAGCCAGCTCGGCTCAGACCTGTGCAACCCCGGGTGGCACCACCATCTA 757
DB 1156 AGAACAGCACCAGCGCCAGCTCAAGGACAACTGACGTCTCTGTTGGGGCCACCATCA 1097
QY 758 TGGACTCCAGCCCTGGAGCAGGGCGGGTGGAGCAGCCACCATGAGCGCGTGGAGCG 817
DB 1096 TGCCTTGATGTGCTGGAGAGTGGGGGCTTCGCTCCCTGCTCATCAGCGCTGTGGAGC 1037
QY 818 TGCCACCTCCGGGGCCAAAGGAGCT 841

DB 1036 CTCCTGCATCCGCACACGGGAGCT 1013

RESULT 12

AAFL5843

ID AAF15843 standard; cDNA; 1478 BP.

XX AAF15843;

AC AAF15843;

XX 13-MAR-2001 (first entry)

DT Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human: prostate cancer; prostate cancer antigen; detection; diagnosis;

KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;

KW vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;

KW antibacterial; gene therapy; neural; immune; reproductive; renal;

KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;

KW wound; infectious disease; ss.

XX Homo sapiens.

OS WO200055174-A1.

PN 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05988.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM;

XX WPI; 2000-587513/55.

XX P-PSDB; AAB56640.

PT Prostate cancer associated gene sequences, referred to as prostate

PT cancer antigens, useful for treatment, prevention, and diagnosis of

PT disorders such as prostate cancer -

XX Claim 1; Page 797-798; 2338pp; English.

XX AAF15566 to AAF16505 encode the human prostate cancer associated

CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.

CC The prostate cancer antigens can have neuroprotective, cytostatic,

CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,

CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,

CC and can be used in gene therapy. The prostate cancer antigen

CC polynucleotides may be used for detection of prostate cancer, chromosome

CC identification, as chromosome markers, and for numerous other diagnostic

CC or research purposes. The prostate cancer antigens may be used to treat

CC disorders such as neural, immune, muscular, reproductive,

CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to

CC AAB57303 represent sequences used in the exemplification of the present

CC invention.

XX Sequence 1478 BP; 292 A; 447 C; 448 G; 288 T; 3 other;

Query Match 15.2%; Score 203.4; DB 21; Length 1478;

Best Local Similarity 57.9%; Pred. No. 6.6e-36;

Matches 466; Conservative 0; Mismatches 316; Indels 23; Gaps 5;

QY 56 GCGCGTGGCTTCGTGGGCGCGCGCATGCGGGGGCCATCGCGCAGGCGCTCATCAG 115

DB 78 GAGCGTGGCTTCATCGCGCGCTGGCCAGCTGGCTTTTCCCTGGCCAAAGGCTTGCACAG 137

QY 116 -AGCAGGAAAAGTGAAGCTCAGCACATCTAGCCAGTGGCCAGTGCACCAACAGGAACTAT 174

DB 138 CAGCAGGCGTCTTGGCTGCCCAAGATATATGGTAGTCTCC---CAGACATGAGCTGG 194

Db 900 GCACCATTCACGCCCTCACTTTCTAGAGAGTGGGGCTTCGCTCTCTGCTCATCAAT 959
Qy 807 GCCGTGGAGGTGCCACCTGCCGGGCCAAGGAGCT 841
Db 960 GCAGTTGAGGCCTCTGTATCCGCAACACGAGAGCT 994

Search completed: July 26, 2003, 05:45:46
Job time : 380 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 03:09:38 ; Search time 3560 Seconds
(without alignments)
10938.077 Million cell updates/sec

Title: US-09-806-536A-29
Perfect score: 1338
Sequence: 1 ggtgagcgagctgttcga.....ggggcagcagcatatggggt 1338

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :					GenEmbl :				
1:	gb_ba:	*	1:	gb_ba:	*	1:	gb_ba:	*	AK023914 Homo sapi
2:	gb_htg:	*	2:	gb_htg:	*	2:	gb_htg:	*	AK001500 Homo sapi
3:	gb_in:	*	3:	gb_in:	*	3:	gb_in:	*	AL833857 Homo sapi
4:	gb_om:	*	4:	gb_om:	*	4:	gb_om:	*	BC007993 Homo sapi
5:	gb_ov:	*	5:	gb_ov:	*	5:	gb_ov:	*	AC105210 Homo sapi
6:	gb_pat:	*	6:	gb_pat:	*	6:	gb_pat:	*	AC087823 Homo sapi
7:	gb_ph:	*	7:	gb_ph:	*	7:	gb_ph:	*	AC067930 Homo sapi
8:	gb_pl:	*	8:	gb_pl:	*	8:	gb_pl:	*	AC019122 Homo sapi
9:	gb_pr:	*	9:	gb_pr:	*	9:	gb_pr:	*	BC026536 Mus muscu
10:	gb_ro:	*	10:	gb_ro:	*	10:	gb_ro:	*	AC120593 Rattus no
11:	gb_sts:	*	11:	gb_sts:	*	11:	gb_sts:	*	BC001504 Homo sapi
12:	gb_sy:	*	12:	gb_sy:	*	12:	gb_sy:	*	AX409754 Sequence
13:	gb_un:	*	13:	gb_un:	*	13:	gb_un:	*	M77836 Human pyro
14:	gb_vi:	*	14:	gb_vi:	*	14:	gb_vi:	*	AF218000 Homo sapi
15:	em_ba:	*	15:	em_ba:	*	15:	em_ba:	*	BC022244 Homo sapi
16:	em_fun:	*	16:	em_fun:	*	16:	em_fun:	*	BC014868 Homo sapi
17:	em_hum:	*	17:	em_hum:	*	17:	em_hum:	*	BC020553 Homo sapi
18:	em_in:	*	18:	em_in:	*	18:	em_in:	*	BC006882 Mus muscu
19:	em_mu:	*	19:	em_mu:	*	19:	em_mu:	*	BC006727 Mus muscu
20:	em_om:	*	20:	em_om:	*	20:	em_om:	*	AF151351 Homo sapi
21:	em_or:	*	21:	em_or:	*	21:	em_or:	*	AL049819 Streptomy
22:	em_ov:	*	22:	em_ov:	*	22:	em_ov:	*	AK001508 Homo sapi
23:	em_pat:	*	23:	em_pat:	*	23:	em_pat:	*	AC116520 Mus muscu
24:	em_ph:	*	24:	em_ph:	*	24:	em_ph:	*	AE013306 Methanosa
25:	em_pl:	*	25:	em_pl:	*	25:	em_pl:	*	AC120830 Rattus no
26:	em_ro:	*	26:	em_ro:	*	26:	em_ro:	*	AC126537 Rattus no
27:	em_sts:	*	27:	em_sts:	*	27:	em_sts:	*	AY071558 Drosophil
28:	em_un:	*	28:	em_un:	*	28:	em_un:	*	AF035580 Methanosa
29:	em_vi:	*	29:	em_vi:	*	29:	em_vi:	*	AE011122 Methanosa
30:	em_htg_hum:	*	30:	em_htg_hum:	*	30:	em_htg_hum:	*	AC009461 Drosophil
31:	em_htg_inv:	*	31:	em_htg_inv:	*	31:	em_htg_inv:	*	AC007804 Drosophil
32:	em_htg_other:	*	32:	em_htg_other:	*	32:	em_htg_other:	*	AE003718 Drosophil
33:	em_htg_mus:	*	33:	em_htg_mus:	*	33:	em_htg_mus:	*	AE012787 Chlorobiu
34:	em_htg_pln:	*	34:	em_htg_pln:	*	34:	em_htg_pln:	*	D25413 Thermus the
35:	em_htg_rtd:	*	35:	em_htg_rtd:	*	35:	em_htg_rtd:	*	AF098020 Drosophil
36:	em_htg_mam:	*	36:	em_htg_mam:	*	36:	em_htg_mam:	*	AF170829 Drosophil
37:	em_htg_vrt:	*	37:	em_htg_vrt:	*	37:	em_htg_vrt:	*	AL646071 Ralstonia
38:	em_sy:	*	38:	em_sy:	*	38:	em_sy:	*	AC017841 Drosophil
39:	em_htgo_hum:	*	39:	em_htgo_hum:	*	39:	em_htgo_hum:	*	AC007891 Drosophil
40:	em_htgo_mus:	*	40:	em_htgo_mus:	*	40:	em_htgo_mus:	*	AC007814 Drosophil
41:	em_htgo_other:	*	41:	em_htgo_other:	*	41:	em_htgo_other:	*	AE003724 Drosophil

10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rtd: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1307	97.7	2331	9	AK023914	Homo sapi
2	1291.8	96.5	2615	9	AK001500	Homo sapi
3	1200.4	89.7	2400	9	HSMB05145	
4	1127.8	84.3	1178	9	BC007993	Homo sapi
5	656.2	49.0	108980	2	AC105210	Homo sapi
6	653.2	48.8	164959	2	AC087823	Homo sapi
7	653.2	48.8	174906	2	AC067930	Homo sapi
8	599.4	44.8	207636	2	AC019122	Homo sapi
9	567.6	42.4	1276	10	BC026536	Mus muscu
10	338.4	25.3	165459	2	AC120593	Rattus no
11	226.4	16.9	1848	9	BC001504	Homo sapi
12	224.8	16.8	1792	6	AX409754	Sequence
13	224.8	16.8	1792	9	HUMP5CR	
14	213.8	16.0	1769	9	AF218000	Homo sapi
15	205.4	15.4	1757	9	BC022244	Homo sapi
16	201.8	15.1	1676	9	BC014868	Homo sapi
17	201.8	15.1	1708	9	BC020553	Homo sapi
18	191.6	14.3	1590	10	BC006882	Mus muscu
19	173.4	13.0	1820	10	BC006727	Mus muscu
20	167	12.5	999	9	AF151351	Homo sapi
21	151.6	11.3	16911	1	SCE7	
22	148.8	11.1	3149	9	AK001508	Homo sapi
23	147	11.0	145544	2	AC116520	Mus muscu
24	143.4	10.7	10490	1	AE013306	Methanosa
25	143.2	10.7	162879	2	AC120830	Rattus no
26	143.2	10.7	196552	2	AC126537	Rattus no
27	141.4	10.6	999	3	AY071558	Drosophil
28	138.4	10.3	5139	1	AF305580	Methanosa
29	138.4	10.3	10357	1	AE011122	Methanosa
30	127.6	9.5	28128	2	AC015301	Drosophil
31	127.6	9.5	171612	3	AC009461	Drosophil
32	127.6	9.5	175440	3	AC007804	Drosophil
33	127.6	9.5	253176	3	AE003718	Drosophil
34	126	9.4	10029	1	AE012787	Chlorobiu
35	123	9.2	2678	1	TTHPROC	
36	116.6	8.7	1089	3	AF098020	Drosophil
37	111.8	8.4	4290	3	AF170829	Drosophil
38	110.4	8.3	203050	1	AL646071	Ralstonia
39	110.2	8.2	45438	2	AC017841	Drosophil
40	110.2	8.2	153048	3	AC007891	Drosophil
41	110.2	8.2	204173	3	AC007814	Drosophil
42	110.2	8.2	232380	3	AE003724	Drosophil
43	107.6	8.0	1017	8	AF466195	Phytophth
44	107	8.0	164959	2	AC087823	Homo sapi
45	105.2	7.9	1356	3	AF241824	Leishmani

ALIGNMENTS

RESULT 1
AK023914
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AK023914
Homo sapiens cDNA FLJ13852 fis, clone THY0100934, weakly similar
to PYRROLINE-5-CARBOXYLATE REDUCTASE (EC 1.5.1.2).
2331 bp mRNA linear PRI 01-AUG-2002
AK023914
GI:10435995
oligo capping; fis (full insert sequence).
Homo sapiens thyroid gland cDNA to mRNA, clone-lib:THY01
clone:THY0100934.

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,

Pred. No. is the number of results predicted by chance to have a

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Isoqai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Ishibashi, T., Fujimori, K.,
Tanai, H., Kinata, M., Watanabe, M., Hiraoka, S., Ishii, S., Kawai, Y.,
Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K.,
Masuho, Y., and Kanehori, K.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2615)
AUTHORS Isoqai, T. and Otsuki, T.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Takao Isoqai, Helix Research Institute,
Genomics Laboratory; 1332-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction: 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
1..2615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP2005690"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP2"
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
BASE COUNT 492 a 831 c 818 g 474 t
ORIGIN
Query Match 96.5%; Score 1291.8; DB 9; Length 2615;
Best Local Similarity 99.1%; Pred. No. 1.7e-221;
Matches 1299; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 27 AAGATGGCAGCTGCGAGCGCTCCGCGGCGCGTTCGTGGGCGCGGCGCGCATG 86
|||||
DB 2615 AAGATGGCAGCTGCGAGCGCTCCGCGGCGCGTTCGTGGGCGCGGCGCGCATG 2556
QY 87 GCGGGGCGCATCGCAGCGGCTCATCAGCAGGAGAAAGTGGAGCTCAGCACAATCTG 146
|||||
DB 2555 GCGGGGCGCATCGCAGCGGCTCATCAGCAGGAGAAAGTGGAGCTCAGCACAATCTG 2496
QY 147 GCCAGTGCCACCAACAGACAGGAACCTATGTCATTTCAGCTCTGGGTGGCGGACACG 206
|||||
DB 2495 GCCAGTGCCACCAACAGACAGGAACCTATGTCATTTCAGCTCTGGGTGGCGGACACG 2436
QY 207 CACTCAACAGGAGGTGTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 266
|||||
DB 2435 CACTCAACAGGAGGTGTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2376
QY 267 GTGCTGCCAGCTGCTCGCAGAGTGGCTCGTGTGTCACACCTGACACATCTTGCTG 326
|||||
DB 2375 GTGCTGCCAGCTGCTCGCAGAGTGGCTCGTGTGTCACACCTGACACATCTTGCTG 2316
QY 327 TCGGTGGCTGCTGGGTGCTCTGAGCACCTGGAGGAGCTGCTGCCCGCCAAACACAGG 386
|||||
DB 2315 TCGGTGGCTGCTGGGTGCTCTGAGCACCTGGAGGAGCTGCTGCCCGCCAAACACAGG 2256
QY 387 GTGCTGCGGCTTGTGCCCACTGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 446
|||||
DB 2255 GTGCTGCGGCTTGTGCCCACTGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 2196
QY 447 CGGGGCGCGCAGTGGGGAGCAGCAGACCAAGCTCTGTCAGCATCTGCTGGAGCGCTGT 506
|||||
DB 2195 CGGGGCGCGCAGTGGGGAGCAGCAGACCAAGCTCTGTCAGCATCTGCTGGAGCGCTGT 2136

QY 507 GGGCGGTGTGAGGAGTGGCTGAAGCCTAGTGCAGATCCACACTGGCCTCAGTGGCAGT 566
|||||
DB 2135 GGGCGGTGTGAGGAGTGGCTGAAGCCTAGTGCAGATCCACACTGGCCTCAGTGGCAGT 2076
QY 567 GGGGTGGCCTTCGTGTGTGTGTCATTTCTCCGAGGCGCTGGCTGAAGAGCCCTCAAGATGGCG 626
|||||
DB 2075 GGGGTGGCCTTCGTGTGTGTGTCATTTCTCCGAGGCGCTGGCTGAAGAGCCCTCAAGATGGCG 2016
QY 627 ATGCCAGCAGCCTGGCCCAACCCATCTGCTGCCCAAGCCTGCTGGGGAGCGCCCAAGATG 686
|||||
DB 2015 ATGCCAGCAGCCTGGCCCAACCCATCTGCTGCCCAAGCCTGCTGGGGAGCGCCCAAGATG 1956
QY 687 CTGCTGCACAGGCGCCAAACACCCAGCCAGCTCGCTGCAGACGTGTGCACCCCGGCTGGC 746
|||||
DB 1955 CTGCTGCACAGGCGCCAAACACCCAGCCAGCTCGCTGCAGACGTGTGCACCCCGGCTGGC 1896
QY 747 ACCACCATCTATGAGCTTCCACGCCCTGGAGCAGGGGGCTGGAGCAGCAGCAGCATGAGC 806
|||||
DB 1895 ACCACCATCTATGAGCTTCCACGCCCTGGAGCAGGGGGCTGGAGCAGCAGCAGCATGAGC 1836
QY 807 GCGGTGGAGCTGCCACCTGCCGGGCCAAGGAGCTCAGCAGAAAGTAGGCTGGGCTGG 866
|||||
DB 1835 GCGGTGGAGCTGCCACCTGCCGGGCCAAGGAGCTCAGCAGAAAGTAGGCTGGGCTGG 1776
QY 867 CCATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 926
|||||
DB 1775 CCATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1716
QY 927 GCTCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 986
|||||
DB 1715 GCTCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1656
QY 987 CAGGGGCGAGGACTTGGGAGGTTCCAGCAGCGGGGAGCCCGAGCAGTGGGGAGCACTC 1046
|||||
DB 1655 CAGGGGCGAGGACTTGGGAGGTTCCAGCAGCGGGGAGCCCGAGCAGTGGGGAGCACTC 1596
QY 1047 CTCCCTCCCGAGTGAGCAGAAAGCAGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1106
|||||
DB 1595 CTCCCTCCCGAGTGAGCAGAAAGCAGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1536
QY 1107 CCAGCTTGTGTCACATTTGTTTGGAGGGGCCCCAAGAGATGGGCTGCTGCTGCTGCTGCTGCTG 1166
DB 1535 CCAGCTTGTGTCACATTTGTTTGGAGGGGCCCCAAGAGATGGGCTGCTGCTGCTGCTGCTGCTG 1476
QY 1167 GCATGGTGGGAGTGGTGGTGGAGCCATGAACAGAACTACGGTAACAGCAGCGGCTGGC 1226
DB 1475 GCATGGTGGGAGTGGTGGAGCCATGAACAGAACTACGGTAACAGCAGCGGCTGGC 1416
QY 1227 CCAATGCTGCTGAGAGCTGGAGCTTGGCTTGGCTTTCAGAGTGGGCTGCTGCTGCTGCTGCTG 1286
DB 1415 CCAATGCTGCTGAGAGCTGGAGCTTGGCTTGGCTTTCAGAGTGGGCTGCTGCTGCTGCTGCTG 1356
QY 1287 CAGCAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1337
DB 1355 CAGCAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1305

RESULT 3
HSM805145 HSM805145 2400 bp mRNA linear PRI 12-JUL-2002
LOCUS Homo sapiens mRNA; cDNA DKF2p761H0716 (from clone DKF2p761H0716).
DEFINITION
ACCESSION AL833857
VERSION 1
KEYWORDS GI:21739340
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2400)
AUTHORS Ansorge, W., Wirkner, U., Mewes, H.W., Well, B. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2002) GSF, Institut fuer Bioinformatik MIPS,

COMMENT
 Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by EMBL (European Molecular Biology Laboratories,
 Heidelberg/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 This clone (DKFZp761H0716) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at <http://mips.gsf.de/proj/cDNA/>.

FEATURES
 source
 Location/Qualifiers
 1..2400
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /clone="DKFZp761H0716"
 /tissue_type="anygdala"
 /clone_lib="761 (synonym: hamy2). Vector pSport1; host
 DH10B; sites NotI + SalI"
 /dev_stage="adult"
 1..2400
 /gene="DKFZp761H0716"
 <1..735
 /gene="DKFZp761H0716"
 /note="similarity to pyroline 5-carboxylate reductase
 isoform (Homo sapiens)"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAD38716.1"
 /db_xref="GI:21739341"
 /translation="GKVEAQHILASAPDRNLCFQALGCRTHNQEVLOSCLLVIF
 EGAIVNARHVGVSSETNLLHLEACRCEVEPAYVDIHTLGSQGVAFVCAFSEA
 LAEGAVMGMPSSLAHRIAQTLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYGLHAL
 EOGGLRAATMSAVEAATCAKELSRK"
 2381
 poly_a_site
 452 a 756 c 740 g 452 t
 /gene="DKFZp761H0716"

BASE COUNT
 ORIGIN
 Query Match 89.7%; Score 1200.4; DB 9; Length 2400;
 Best Local Similarity 99.1%; Pred. No. 3.9e-205;
 Matches 1207; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 120 GGAAGAGTGGAGCTCAGACATATCTGGCCAGTCACCAACAGAGAACTATGTAC 179
 DB 1 GGAAGAGTGGAGCTCAGACATATCTGGCCAGTCACCAACAGAGAACTATGTAC 60
 QY 180 TTTCAGCTCTGGGTTGCCGACACGCTCATGTGCTGCCAGTGTCTCTGAGAGTGCCTG 239
 DB 61 TTTCAGCTCTGGGTTGCCGACACGCTCATGTGCTGCCAGTGTCTCTGAGAGTGCCTG 120
 QY 240 CTCGTCATCTTTGCCACCAAGCTCATGTGCTGCCAGTGTCTCTGAGAGTGCCTCT 299
 DB 121 CTCGTCATCTTTGCCACCAAGCTCATGTGCTGCCAGTGTCTCTGAGAGTGCCTCT 180
 QY 300 GTGGTCCACCTGAACACATCTTGTGTCCTGGTGTCTGAGAGTGCCTCTGAGACCCCTG 359
 DB 181 GTGGTCCACCTGAACACATCTTGTGTCCTGGTGTCTGAGAGTGCCTCTGAGACCCCTG 240
 QY 360 GAGGAGCTCTGCCGCCCAACACAGGCTGTGTCGGGTCTGCGGCTGTCTGAGAGTGCCTG 419
 DB 241 GAGGAGCTCTGCCGCCCAACACAGGCTGTGTCGGGTCTGCGGCTGTCTGAGAGTGCCTG 300
 QY 420 GTCCAGGAGGGCCATAGTATGTCGCGCGGCGCCACGCTGGGAGCAGGAGACCAAG 479
 DB 301 GTCCAGGAGGGCCATAGTATGTCGCGCGGCGCCACGCTGGGAGCAGGAGACCAAG 360
 QY 480 CTCCTCCACATCTCTGGAGCCCTGTGGGGGGTGTGAGAGTGCCTGAGAGTGCCTG 539
 DB 361 CTCCTCCACATCTCTGGAGCCCTGTGGGGGGTGTGAGAGTGCCTGAGAGTGCCTG 420

QY 540 GACATCCACACTGGCCTCAGTGGCAGTGGCTGGCGCTTCGTGTGTGATCTTCGAGGCC 599
 DB 421 GACATCCACACTGGCCTCAGTGGCAGTGGCTGGCGCTTCGTGTGTGATCTTCGAGGCC 480
 QY 600 CTGGCTGAAGGAGCGCTCAAGATGGGATGCCAGAGCGCTGGCCACCGCATCGCTGCC 659
 DB 481 CTGGCTGAAGGAGCGCTCAAGATGGGATGCCAGAGCGCTGGCCACCGCATCGCTGCC 540
 QY 660 CAGACCCCTGCTGGGAGCGCCAAAGATGCTGTGACGAGGGCCAAACACCGCCAGCTG 719
 DB 541 CAGACCCCTGCTGGGAGCGCCAAAGATGCTGTGACGAGGGCCAAACACCGCCAGCTG 600
 QY 720 CGCTCAGACGTGTGCACCCCGGGTGGCACCACCATCTATGACATCCACGCGCTGGAGCAG 779
 DB 601 CGCTCAGACGTGTGCACCCCGGGTGGCACCACCATCTATGACATCCACGCGCTGGAGCAG 660
 QY 780 GCGGGCTCCAGCAGCGCCACCATGAGCGCGCTGAGGCTGCCACTGCCGGGCCAAGGAG 839
 DB 661 GCGGGCTCCAGCAGCGCCACCATGAGCGCGCTGAGGCTGCCACTGCCGGGCCAAGGAG 720
 QY 840 CTCAGCAGAAAGTAGGCTGGGCTCTGGCCTATCCCTTCTGCTGCTGTGCCCCCTGCTC 899
 DB 721 CTCAGCAGAAAGTAGGCTGGGCTCTGGCCTATCCCTTCTGCTGCTGTGCCCCCTGCTC 780
 QY 900 CCTGTGTCCTCCCTCCCTCAGGACTGGCGCTCCCTCCCTCCTGCAATGAGGCTCTCTACTG 959
 DB 781 CCTGTGTCCTCCCTCAGGACTGGCGCTCCCTCCCTCCTGCAATGAGGCTCTCTACTG 840
 QY 960 CTCCTTCTCCCTTGCACAGGAAATGCAGGGGCGAGGACTTGGAGGTTCCAGCAGCGG 1019
 DB 841 CTCCTTCTCCCTTGCACAGGAAATGCAGGGGCGAGGACTTGGAGGTTCCAGCAGCGG 900
 QY 1020 GGGGAGCGCCGACCACTCTCTCCCTCCCTCCCTCCCTGCAATGAGGCTCTCTACTG 1079
 DB 901 GGGGAGCGCCGACCACTCTCTCCCTCCCTCCCTCCCTGCAATGAGGCTCTCTACTG 960
 QY 1080 GTGCTCTGCCCTTGTGTCAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1139
 DB 961 GTGCTCTGCCCTTGTGTCAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
 QY 1140 AAGAGATGGCTCTTGTGTCATTTGCCCGCATGTTGGCAGTGGTTGAGGCCATGAACA 1199
 DB 1021 AAGAGATGGCTCTTGTGTCATTTGCCCGCATGTTGGCAGTGGTTGAGGCCATGAACA 1080
 QY 1200 GAATTTACGTTAAGCAGCAGCGCTGGCCCAATGCTGCTGGAGCTGGAGCTTGGCTTT 1259
 DB 1081 GAATTTACGTTAAGCAGCAGCGCTGGCCCAATGCTGCTGGAGCTGGAGCTTGGCTTT 1140
 QY 1260 GGCTTTCCAAAGTGGGCTGTCAGCTACAGCCAGCGCGGCGGCGGCGGCGGCGGCGG 1319
 DB 1141 GGCTTTCCAAAGTGGGCTGTCAGCTACAGCCAGCGGCGGCGGCGGCGGCGGCGGCGG 1260
 QY 1320 GGGCAGCAGCATATGGG 1337
 DB 1201 GGGCAGCAGCATATGGG 1218

RESULT 4
 BC007993
 LOCUS
 DEFINITION
 Homo sapiens, hypothetical protein FLJ13852, clone MGC:14985
 IMAGE:3533609, mRNA, complete cds.
 ACCESSION
 BC007993
 VERSION
 BC007993.1
 KEYWORDS
 MGC.
 SOURCE
 Homo sapiens.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 1178)
 AUTHORS
 Strausberg,R.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (15-MAY-2001) National Institutes of Health, Mammalian

REFERENCE
AUTHORS

2 (bases 1 to 108980)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Submitted (26-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS

3 (bases 1 to 108980)
 Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 5, 2002 this sequence version replaced gi:21307003.

COMMENT

All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L23062
 Center clone name: 118_P_12

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 14262: contig of 14262 bp in length
 * 14263 14362: gap of 100 bp
 * 14363 16379: contig of 2017 bp in length

* 16380	16479:	gap of	100 bp
* 16480	18932:	contig of 2453 bp in length	
* 18933	19032:	gap of	100 bp
* 19033	25980:	contig of 6948 bp in length	
* 25981	26080:	gap of	100 bp
* 26081	47787:	contig of 21707 bp in length	
* 47788	47887:	gap of	100 bp
* 47888	82335:	contig of 34448 bp in length	
* 82336	82435:	gap of	100 bp
* 82436	97616:	contig of 15181 bp in length	
* 97617	97716:	gap of	100 bp
* 100357	100456:	contig of 2640 bp in length	
* 100457	105947:	contig of 5491 bp in length	
* 105948	106047:	gap of	100 bp
* 106048	108980:	contig of 2933 bp in length.	

FEATURES
 source
 1. 108980
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-118P12"
 /clone_lib="RPC1-11 Human Male BAC"
 BASE COUNT 24096 a 30961 c 30570 g 22231 t 1122 others
 ORIGIN

Query Match	49.0%	Score 656.2;	DB 2;	Length 108980;
Best Local Similarity	93.9%	Pred. No. 4.1e-108;		
Matches 682;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0;

QY	612	GCCCTCAAGATGGGCATGCCAGCAGCGCTGGCCACCGCATCGCTGCCAGACCCCTGCTG	671
Db	56072	GGCCACAGTGGGGACCGCACCACCTCCCGCTCACAGGCCCTTAACGCTCTCCTCAG	56013
QY	672	GGGACGCCAAGATGCTGCTGCAGGAGGGCCACACCCAGCCAGCTGCGCTCAGAGGTG	731
Db	56012	GGGACGCCAAGATGCTGCTGCAGGAGGGCCACACCCAGCCAGCTGCGCTCAGAGGTG	55953
QY	732	TGACACCCCGGGTGGCCACCACTATATGACTCCAGCCCTGGAGCAGGGCGGCTGCCGA	791
Db	55952	TGCACCCCGGGTGGCCACCACTATATGACTCCAGCCCTGGAGCAGGGCGGCTGCCGA	55893
QY	792	GCAGCCACCATGAGCGCGGTGGAGGTGCGACCTGCGGGCCAGGAGCTCAGCAAAAG	851
Db	55892	GCAGCCACCATGAGCGCGGTGGAGGTGCGACCTGCGGGCCAGGAGCTCAGCAAAAG	55833
QY	852	TAGCTGGGCTTGGCCATCTTCTGCTCTGTGCCCCCTGCTCTCCCTGTGTCCTT	911
Db	55832	TAGCTGGGCTTGGCCATCTTCTGCTCTGTGCCCCCTGCTCTCCCTGTGTCCTT	55773
QY	912	CCCTGAGGACTGCGGCTCCCTCCCTGCTCATGAGGTCTCTACTGCTCTCTCCCT	971
Db	55772	CCCTGAGGACTGCGGCTCCCTCCCTGCTCATGAGGTCTCTACTGCTCTCTCCCT	55713
QY	972	TTGCACAGGAAATGAGGGGCGAGGACTTGGAGGTTCACGACGCGGGGAGCCCCGA	1031
Db	55712	TTGCACAGGAAATGAGGGGCGAGGACTTGGAGGTTCACGACGCGGGGAGCCCCGA	55653
QY	1032	CCAGTGGGACACTCTCTCCCTCCAGTGAGCAGAGGACCGTGGTGGCTCTGCC	1091
Db	55652	CCAGTGGGACACTCTCTCCCTCCAGTGAGCAGAGGACCGTGGTGGCTCTGCC	55593
QY	1092	CTTGCTGCAGTGAGCCACCTTCTGCTCAACATTTGTTCTGAGGGGCCCAAGAGATGGCCT	1151
Db	55592	CTTGCTGCAGTGAGCCACCTTCTGCTCAACATTTGTTCTGAGGGGCCCAAGAGATGGCCT	55533
QY	1152	CTTGCTCAATTTGCCCGCATGGTTGGGCGAGTTGGTTGAGGCCATGAACAGAACTTACGTA	1211
Db	55532	CTTGCTCAATTTGCCCGCATGGTTGGGCGAGTTGGTTGAGGCCATGAACAGAACTTACGTA	55473
QY	1212	ACAGGCAGCGCTGGCCCAATGCTGCTGGAGCTGGAGCTTGCCTTTGCCCTTCCAACT	1271

```
Db      55472 ACAGGCACGGCTGCCCAATGCTGGTCTCGAGCTGGAGCTGTCCTTTGGCTTTCCAGGT 55413
QY      1272 GGGCTCTGTCAGCTACAGCCAGCGCGCTGCCTCATCTCAGCTCTAGGGGGGACGACCAT 1331
        |||
Db      55412 GGCCTCCGTGACGTACAGCCAGCGCGCTGCCTCATCTCAGCTCTAGGGGGGACGACCCA 55353
QY      1332 ATGGGG 1337
        |||
Db      55352 TATGGG 55347

RESULT 6
AC087823/c
LOCUS   AC087823               164959 bp    DNA        linear      HTG 16-MAR-2001
DEFINITION Homo sapiens chromosome 8 clone RP11-545C16 map 8, *** SEQUENCING
AC087823
VERSION AC087823.2 GI:13357313
KEYWORDS HTG; HTGS_PHASE1.
SOURCE  Homo sapiens.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 164959)
        Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE    Homo sapiens chromosome 8, clone RP11-545C16
JOURNAL  Unpublished
AUTHORS  2 (bases 1 to 164959)
        Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
        Barre,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
        Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
        Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
        Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
        Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
        Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
        Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
        Lehocsky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
        Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
        McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
        Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
        O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
        Phunkhang,P., Pleyre,N., Pollara,V., Raymond,C., Retta,R.,
        Riback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
        Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
        Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
        Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
        Trauers,M., Travis,N., Triglio,J., Vassiliev,H., Viel,R., Vo,A.,
        Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
        Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2001 this sequence version replaced gi:12584321.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1NR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11736
Center clone name: 545_C_16
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```


* 55532 55631: gap of 100 bp
* 55632 57290: contig of 1659 bp in length
* 57291 57390: gap of 100 bp
* 57391 58997: contig of 1607 bp in length
* 58998 59097: gap of 100 bp
* 59098 60543: contig of 1446 bp in length
* 60544 60643: gap of 100 bp
* 60644 62329: contig of 1686 bp in length
* 62330 62429: gap of 100 bp
* 62430 64069: contig of 1640 bp in length
* 64070 64169: gap of 100 bp
* 64170 66375: contig of 2206 bp in length
* 66376 66475: gap of 100 bp
* 66476 68972: contig of 2497 bp in length
* 68973 69072: gap of 100 bp
* 69073 70628: contig of 1556 bp in length
* 70629 70728: gap of 100 bp
* 70729 73435: contig of 2707 bp in length
* 73436 73535: gap of 100 bp
* 73536 75771: contig of 2236 bp in length
* 75772 75871: gap of 100 bp
* 75872 78176: contig of 2305 bp in length
* 78177 78276: gap of 100 bp
* 78277 80144: contig of 1868 bp in length
* 80145 80244: gap of 100 bp
* 80245 82363: contig of 2119 bp in length
* 82364 82463: gap of 100 bp
* 82464 84226: contig of 1763 bp in length
* 84227 84326: gap of 100 bp
* 84327 87431: contig of 3105 bp in length
* 87432 87531: gap of 100 bp
* 87532 90276: contig of 2745 bp in length
* 90277 90376: gap of 100 bp
* 90377 93023: contig of 2647 bp in length
* 93024 93123: gap of 100 bp
* 93124 95831: contig of 2708 bp in length
* 95832 95931: gap of 100 bp
* 95932 99011: contig of 3080 bp in length
* 99012 99111: gap of 100 bp
* 99112 101722: contig of 2611 bp in length
* 101723 101822: gap of 100 bp
* 101823 105034: contig of 3212 bp in length
* 105035 105134: gap of 100 bp
* 105135 108113: contig of 2979 bp in length
* 108114 108213: gap of 100 bp
* 108214 112105: contig of 3892 bp in length
* 112106 112205: gap of 100 bp
* 112206 115406: contig of 3201 bp in length
* 115407 115506: gap of 100 bp
* 115507 119714: contig of 4208 bp in length
* 119715 119814: gap of 100 bp
* 119815 125158: contig of 5344 bp in length
* 125159 125258: gap of 100 bp
* 125259 130255: contig of 4997 bp in length
* 130256 130355: gap of 100 bp
* 130356 135663: contig of 5308 bp in length
* 135664 135763: gap of 100 bp
* 135764 141345: contig of 5582 bp in length
* 141346 141445: gap of 100 bp
* 141446 146381: contig of 4936 bp in length
* 146382 146481: gap of 100 bp
* 146482 155041: contig of 8560 bp in length
* 155042 155141: gap of 100 bp
* 155142 164370: contig of 9229 bp in length
* 164371 164470: gap of 100 bp
* 164471 164959: contig of 489 bp in length.
Location/Qualifiers

FEATURES

Query Match 48.8%; Score 653.2; DB 2; Length 164959;
Best Local Similarity 98.1%; Pred. No. 1.3e-107;
Matches 661; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
664 CCCTGCTGGGACGCCAAGATGCTGCTGCACGAGGGCCACACCCAGCTGCCT 723

Db 141067 CTCCTCAGGGGACGGCCAAAGATGCTGCTGCACGAGGGCCAAACCCAGCCAGCTGCCT 141008
Qy 724 CAGACGTGTGCACCCCGGTTGGCACCACCATCTATGGACTCCACGCCCTGGAGCGGGC 783
Db 141007 CAGACGTGTGCACCCCGGTTGGCACCACCATCTATGGACTCCACGCCCTGGAGCGGGC 140948
Qy 784 GGCTGCAGCAGCCACCACCATGAGCGCCGTGGAGGCTGCACCTCCCGGGCCAAAGAGCTCA 843
Db 140947 GGCTGCAGCAGCCACCACCATGAGCGCCGTGGAGGCTGCACCTCCCGGGCCAAAGAGCTCA 140888
Qy 844 GCAGAAAGTAGGCTGGGCTCTGGCCATCCCTTCTGCTGCTCTGCCCCCTGCTCTCCCTG 903
Db 140887 GCAGAAAGTAGGCTGGGCTCTGGCCATCCCTTCTGCTGCTCTGCCCCCTGCTCTCCCTG 140828
Qy 904 TGTCCCTTCCCTGAGGACTCGGCTCCCTCCCTCCCTGCTGATGAGGCTCTCTACTGCTCC 963
Db 140827 TGTCCCTTCCCTGAGGACTCGGCTCCCTCCCTCCCTGCTGATGAGGCTCTCTACTGCTCC 140768
Qy 964 TTTCCCTTTCACAGGGAATGAGGGGCGAGACTTTGGAGGTTCCAGCAGCGGGGG 1023
Db 140767 TTTCCCTTTCACAGGGAATGAGGGGCGAGACTTTGGAGGTTCCAGCAGCGGGGG 140708
Qy 1024 AGCCCGACAGTGGGGACACTCTCCCTCCCGAGTGAGCAGAGGCACCGTGGTGG 1083
Db 140707 AGCCCGACAGTGGGGACACTCTCCCTCCCGAGTGAGCAGAGGCACCGTGGTGG 140648
Qy 1084 CTCCTGCCCTTGTGTCAGTGAGCCACCTTGTCTGCAACATTGGTTCTGAGGGGCCCAAGA 1143
Db 140647 CTCCTGCCCTTGTGTCAGTGAGCCACCTTGTCTGCAACATTGGTTCTGAGGGGCCCAAGA 140588
Qy 1144 GATGGCTCTTGTTCATTTGCCGCATGTTGGCAGTTGGGAGTTCCAGCAGCGGGGG 1203
Db 140587 GATGGCTCTTGTTCATTTGCCGCATGTTGGCAGTTGGGAGTTCCAGCAGCGGGGG 140528
Qy 1204 TTACGGTAACAGGCACGGCTGGCCCAATGCTGCTGAGAGTGGAGCTTGCCTTTGGCT 1263
Db 140527 TTACGGTAACAGGCACGGCTGGCCCAATGCTGCTGAGAGTGGAGCTTGCCTTTGGCT 140468
Qy 1264 TTCCAAGTGGGCTGTCAGCTACAGCCAGCGGGCTGCCTCATCTCAGCTCAGCTAGGGGG 1323
Db 140467 TTCCAAGTGGGCTGTCAGCTACAGCCAGCGGGCTGCCTCATCTCAGCTCAGCTAGGGGG 140408
Qy 1324 ACGAGCATATGGGG 1337
Db 140407 ACGAGCATATGGGG 140394

RESULT 7

AC067930/c AC067930 174906 bp DNA linear PRI 24-JUL-2002
LOCUS Homo sapiens chromosome 8, clone RP11-661A12, complete sequence.
DEFINITION AC067930
ACCESSION AC067930
VERSION AC067930.7 GI:21954034
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174906)

Birren,B., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 8, clone RP11-661A12

Unpublished

2 (bases 1 to 174906)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boquelavkiy,L., Boukhalter,B., Brown,A., Burkett,G.,

Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (27-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 174906)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (21-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 174906)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (24-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 24, 2002 this sequence version replaced gi:21490265.

All repeats were identified using RepeatMasker:

Smit, A.P.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

FEATURES	Source	Location/Qualifiers	Center project name: LI0100	Center clone name: 661_A_12
repeat_region	1. 174906	/organism="Homo sapiens"		
repeat_region		/db_xref="taxon:9606"		
repeat_region		/chromosome="8"		
repeat_region		/map="8"		
repeat_region		/clone="RP11-661A12"		
repeat_region		/clone_lib="RPC1-11 Human Male BAC"		
repeat_region		complement(351..430)		
repeat_region		/rpt_family="MIR"		
repeat_region		977..1108		
repeat_region		/rpt_family="L1MC4a"		
repeat_region		complement(1149..1449)		
repeat_region		/rpt_family="AluSq"		
repeat_region		3675..3781		
repeat_region		/rpt_family="MIR"		
repeat_region		complement(16671..16974)		
repeat_region		/rpt_family="AluSq"		
repeat_region		20174..20412		
repeat_region		/rpt_family="L2"		
repeat_region		20487..20585		
repeat_region		/rpt_family="L2"		
repeat_region		complement(20694..20702)		
repeat_region		/note="<30 qual SINGL region"		
repeat_region		23655..23772		
repeat_region		/rpt_family="L1ME2"		
repeat_region		24264..24383		
repeat_region		/rpt_family="L1ME2"		
repeat_region		complement(24446..24779)		
repeat_region		/rpt_family="MLT1B"		
repeat_region		26300		
repeat_region		/note="probably A; possibly G"		
repeat_region		26681..26825		
repeat_region		/rpt_family="L1ME2"		
repeat_region		28798..29091		
repeat_region		/rpt_family="AluX"		
repeat_region		32096..32118		
repeat_region		/rpt_family="AT-rich"		
repeat_region		32990..33157		
repeat_region		/rpt_family="L1M2"		
repeat_region		33598..33749		
repeat_region		/rpt_family="L1ME2"		
repeat_region		complement(33750..34064)		
repeat_region		/rpt_family="AluJo"		
repeat_region		34065..34129		
repeat_region		/rpt_family="L1ME2"		
repeat_region		34132..34706		
repeat_region		/rpt_family="L1MC4a"		
repeat_region		34991..35084		
repeat_region		/rpt_family="L1MD3"		
repeat_region		35191..35325		
repeat_region		/rpt_family="L1MC4"		
repeat_region		36901..37069		
repeat_region		/rpt_family="CGGGG)n"		
repeat_region		37077..37162		
repeat_region		/rpt_family="CGGGG)n"		
repeat_region		complement(38672..38844)		
repeat_region		/rpt_family="MSTC"		
repeat_region		complement(38859..38884)		
repeat_region		/note="<30 qual SINGL region"		
repeat_region		complement(38898..38902)		
repeat_region		/note="probably ACTGA"		
repeat_region		complement(38984..39281)		
repeat_region		/rpt_family="AluX"		
repeat_region		complement(39282..39334)		
repeat_region		/rpt_family="MSTC"		

```
repeat_region complement(39366..39668)
repeat_region /rpt_family="AluJb"
repeat_region 40044..40423
repeat_region /rpt_family="MLT1A2"
repeat_region 41752..41933
repeat_region /rpt_family="(CA)n"
repeat_region 42478..42510
repeat_region /rpt_family="G-rich"
unsure complement(43664..43676)
unsure /note="<30 qual SNGL region"
unsure complement(43666..43674)
unsure /note="probably ACTGGTGGC"
unsure 46481..46483
unsure /note="probably CCC"
repeat_region 54691..54749
repeat_region /rpt_family="(CA)n"
repeat_region 56353..56388
repeat_region /rpt_family="MER5B"
repeat_region 60486..60780
repeat_region /rpt_family="AluSx"
repeat_region 60802..61091
repeat_region /rpt_family="AluJb"
repeat_region complement(69122..69423)
repeat_region /rpt_family="AluJb"
repeat_region 71273..71458
repeat_region /rpt_family="L2"
repeat_region 71493..71733
repeat_region /rpt_family="L2"
repeat_region complement(71849..72140)
repeat_region /rpt_family="AluSx"
repeat_region 72167..72461

Query Match 48.8%; Score 653.2; DB 9; Length 174906;
Best Local Similarity 98.1%; Pred. No. 1.3e-107;
Matches 661; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 664 CCTGCTGGGACGCCAAGATGCTGTCAGAGGGCCACACCCAGCCAGCTGGCGT 723
Db 144055 CTCCTCAGGGGACGGCCAGAGATGCTGTCAGAGGGCCACACCCAGCCAGCTGGCGT 143996

QY 724 CAGAGGTGTGACCCCGGGTGGCCACCACCATCTATGGACTCCAGCCCTGGAGCAGGGCG 783
Db 143995 CAGAGGTGTGACCCCGGGTGGCCACCACCATCTATGGACTCCAGCCCTGGAGCAGGGCG 143936

QY 784 GCCTGGGAGCACCACCATGAGCGCGTGGAGGTGCCACCTGCCGGGCCAAGGAGCTCA 843
Db 143935 GCCTGGGAGCACCACCATGAGCGCGTGGAGGTGCCACCTGCCGGGCCAAGGAGCTCA 143876

QY 844 GCAGAAAGTAGGCTGGGCTGGCCATCTTTCTGCTCTGTGCCCTGCTCTCCCTG 903
Db 143875 GCAGAAAGTAGGCTGGGCTGGCCATCTTTCTGCTCTGTGCCCTGCTCTCCCTG 143816

QY 904 TGTCCCTTCCCTGAGGACTCGGCTCCCTCCCTCCCTGTCATGAGGGTCTCTACTGCTCC 963
Db 143815 TGTCCCTTCCCTGAGGACTCGGCTCCCTCCCTCCCTGTCATGAGGGTCTCTACTGCTCC 143756

QY 964 TTCTCCCTTGCACAGGAAATGACAGGGGCGAGACTTGGAGGTTCACAGCAGCGGGG 1023
Db 143755 TTCTCCCTTGCACAGGAAATGACAGGGGCGAGACTTGGAGGTTCACAGCAGCGGGG 143696

QY 1024 AGCCCCGACCACTAGTGGGACACTCTCTCCCTCCCTGAGCAGAGGACCGTGTGTGG 1083
Db 143695 AGCCCCGACCACTAGTGGGACACTCTCTCCCTCCCTGAGCAGAGGACCGTGTGTGG 143636

QY 1084 CTCTGCCCTTCTGTCAGTGAAGCCACCTTGTCTCAACATGTTCTGAGGGGCCCAAGA 1143
Db 143635 CTCTGCCCTTCTGTCAGTGAAGCCACCTTGTCTCAACATGTTCTGAGGGGCCCAAGA 143576

QY 1144 GATGGCGTCTGGTCATTTGGCCCGCATGTTGGCAGTTGTTGAGGCCATGAACAAC 1203
Db 143575 GATGGCGTCTGGTCATTTGGCCCGCATGTTGGCAGTTGTTGAGGCCATGAACAAC 143516

QY 1204 TTACGGTAACAGGCACGGCTGGCCCAATGCTGTCTGGAGCTGGAGCTTGCTTTGGCT 1263
```

```
Db 143515 TTACGGTAACAGGCACGGCTGGCCCAATGCTGTGGAGCTGGAGCTTGCTTTGGCT 143456
QY 1264 TTCCAAGTGGCTCGTCGACGTACAGCCAGCGCGCTGCTCATCTCAGCTCTAGGGGCG 1323
Db 143455 TTCCAGTGGCTCGTCGACGTACAGCCAGCGCGCTGCTCATCTCAGCTCTAGGGGCG 143396
QY 1324 ACAGCATATGGGG 1337
Db 143395 ACAGCATATGGG 143382

RESULT 8
AC019122 207636 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 8 clone RP11-545C16, WORKING DRAFT
DEFINITION SEQUENCE, 12 unordered pieces.
ACCESSION AC019122.3 GI:8318560
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 207636)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207636)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 7, 2000 this sequence version replaced gi:7109639.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: LNH0545C16
----- Summary Statistics -----
Sequencing vector: M13; 88%
Sequencing vector: plasmid; 12%
Chemistry: Dye-primer ET; 88% of reads
Chemistry: Dye-terminator Big Dye; 12% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 198432 bases at least Q40
Consensus quality: 201424 bases at least Q30
Consensus quality: 203070 bases at least Q20
Insert size: 182000; agarose-fp
Quality coverage: 6.36 in Q20 bases; agarose-fp
Quality coverage: 5.66 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1653: contig of 1653 bp in length
* 1654 1753: gap of unknown length
* 1754 3687: contig of 1934 bp in length
* 3688 3787: gap of unknown length
* 3788 11012: contig of 7225 bp in length
* 11013 11112: gap of unknown length
* 11113 16540: contig of 5428 bp in length
* 16541 16640: gap of unknown length
* 16641 23739: contig of 7099 bp in length
```

```

/imap=FVB/N
/clone=MGC:35
/tissue_type="

```

```

/map="FVB/N"
/clone="MGC:35759 IMAGE:4987954"

```

```

/map="FVB/N"
/clone="MGC:35759 IMAGE:4987954"

```

```
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
16 . 840
/codon_start=1
/product="RIKEN cDNA 1110058B13 gene"
/protein_id="AAH26536.1"
/db_xref="GI:20071634"
/db_xref="LocusID:66194"
/translation="MAATMSPRRRVFGVAGRMATATARGLIQAGKVEAKOVLASAPT
DNILCFRALGCPHTNSHNEVLQNCPLVIFATKQVLPVLAENVAPITVEHIIIVSA
AGISLSTMEGLLPNTRVLRVSPNCPVQBGAMARGHAGNAGHAGNAGHAGNAGHAGN
ACIEVPSYVDHILSGSVAFNFTSEALAEAGAIKMGPSGLAHRAAQTILGTAK
MLQQRKHPAQLRTDVLTPAGTTHGLHALERGGFRATMSAVEAATCRKELSKK"
BASE COUNT 288 a 357 c 360 g 271 t
ORIGIN
Query Match 42.4%; Score 567.6; DB 10; Length 1276;
Best Local Similarity 79.8%; Pred. No. 7.2e-92;
Matches 669; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 24 AACAAATGATGGCAGCTGCGGAGCGCTCCGGCGCGCTGCTGCGGCGCGGCGGC 83
DB 10 ACAGATGATGGCGGCGACATGCTGAACCTGCGGCTGCTGCGGCTGCTGCGGCGCGC 69
QY 84 ATGGCGGGGCGCATCGCGAGGCGCTCATCATGAGCAGGAGAAAGTGGAGCTCAGCACA 143
DB 70 ATGGCGGGGCGCATGCGCGAGGCGCTCATCAGCAGGAGGAGGAGGAGGAGGAGGAGG 129
QY 144 CTGGCCAGTGCACCAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 203
DB 130 CTGGCCAGTGCACCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 189
QY 204 AGCACTCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 263
DB 190 ACTCACTCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 249
QY 264 CATGTGCTGCCAGCTGCTGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 323
DB 250 CAAGTCTGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 309
QY 324 GTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 383
DB 310 GTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 369
QY 384 CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 443
DB 370 CGAGTATTCGAGAGTCTCTCCCAATCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 429
QY 444 GCGCGGGCGCGCAGCTGCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 503
DB 430 GCGCGGGCGCGCAGCTGCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 489
QY 504 TGTGCGCGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 563
DB 490 TGTGCGCGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 549
QY 564 AGTGCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 623
DB 550 AGTGCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 609
QY 624 GGCATGCCAGCAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 683
DB 610 GGCATGCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 669
QY 684 ATGCTGCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 743
DB 670 ATGCTGCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 729
QY 744 GGCACCCACCATCTAGGAGTCCAGGCTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 803
DB 730 GGAACCCACCATCTAGGAGTCCAGGCTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 789
```

```
QY 804 AGCGCGCTGTGAGGCTGCCACCTGCGCGGCGCGGAGGAGCTCAGCAAGAGTAGGCTGGC 861
DB 790 AGTGCCTGGAAGCAGCTACCTGCGGCTGAGGAGCTCAGCAAGAGTAGGAGGAGCAGC 847

RESULT 10
AC120593 165459 bp DNA linear HTG 24-AUG-2002
LOCUS Rattus norvegicus clone CH230-300L19, *** SEQUENCING IN PROGRESS
DEFINITION ** 58 unordered pieces.
AC120593
VERSION AC120593.3 GI:22164892
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 165459)
Muzny,D,Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angulano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,J., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Devila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hayes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Huiyik,S., Hume,J., Iolebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Loulsegged,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., Mcleod,M., McNeill,T., Meenen,E., Milosavljevic,A.,
Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwaokemeleh,O., Okwuonu,G.,
Olarpunsgoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A.,
Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Sisson,I., Sitter,C.D., Smajk,D., Sneed,A., Sodergren,E.,
Song,X.-Z., Sorrelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
Syatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
Waldron,L., Walker,B., Wang,O., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlaczky,R.,
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
Yen,J., Yoon,L., Yoon,F., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
Direct Submission
TITLE
JOURNAL
REFERENCE 2 (bases 1 to 165459)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
```



```
QY 518 GGAGGTGCTGAAGCCTACGTCGACATCCACACTGCGCCTCAGTGGCAGTGGCGTGCCCTT 577
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 GGAGGTGGAAGGACCTGATGTGATGCGCTCAGCGGGCTCAGTGGCAGGCGCCGCCTA 588
QY 578 CGTGTGTGATTTCTCGAGGCCCTGGCTGAAGAGCCGCTCAAGATGGGCGTCCCGCAG 637
|| || || || || || || || || || || || || || || || || || || || || ||
Db 589 CGCATTCACAGCCCTGATGCTGCTGATGGGGCGTGAAGATGGGACTTCCAAGGCG 648
QY 638 CTGCGCCACCGCATCGCTGCCAGACCCTGCTGGGAGCGGCCAAGATGCTGCTGCAGA 697
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 649 CTGGCAGTCCGCTCGGGGCCAGGCCCTCTCGGGGCTGCCAAGATGCTGCTGCATC 708
QY 698 GGSCCAACACCCAGCCAGCTGCGCTCAGACGTGTGCACCCCGGTGGCACCACCATCTA 757
|| || || || || || || || || || || || || || || || || || || || || ||
Db 709 AGAACAGCACCAGGCCACCTCAAGGACAACGTCAAGTCTCTGCTGGGGCCACCATCA 768
QY 758 TGGACTCCAGCCCTGGAGCAGCGGGCTGGCAGCAGCCACCATGAGCGCGTGGAGC 817
|| || || || || || || || || || || || || || || || || || || || || ||
Db 769 TGCCTTGCAATGCTGGAGAGTGGGGCTTCGGCTCCCTGCTCATCAACAGCTGTGGAGC 828
QY 818 TGCCACCTGCGGGCCAAAGGACT 841
Db 829 CTCCTGCATCCGACACGGGAGCT 852

RESULT 12
AX409754
LOCUS AX409754 1792 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 2401 from Patent WO0229103.
ACCESSION AX409754
VERSION AX409754.1 GI:21442459
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 2401 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
source Location/Qualifiers
1..1792
/organism="Homo sapiens"
/db_xref="taxon:9606"
/notes="ENBL/GenBank Accession No. M77836"
BASE COUNT 342 a 573 c 522 g 355 t
ORIGIN

Query Match 16.8%; Score 224.8; DB 6; Length 1792;
Best Local Similarity 58.0%; Pred. No. 1.4e-30;
Matches 466; Conservative 0; Mismatches 317; Indels 21; Gaps 3;

QY 56 GCGCGTGGGCTTCGTGGGCGCGCGCATGGCGGGGCCATCGCGCAGGGCCTCATCAG 115
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14 GAGCGTGGGCTTCATCGCGCTGGCCAGCTGCTTTCGCCCTGGCCAAAGGCTTCACAGC 73
QY 116 AGCAGGAAAGTGGAGCTCAGCACAATCTGCGCAGTGCACCAACAGCAGGAACCTATG 175
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 74 AGCAGGCGTCTTGGCTGCCCAAGATAATGGCTAGCTCC---CAGACATGGACTGGC 130
QY 176 TCACCTTTCAAGCTC-----TGGGTTGCCGACCGCACTCCACCCAGGAGTGCT 226
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 131 CACAGTTTCTGCTCTCAGGAAGATGGGGTGAAGTTGACACCCCAACAGGACGCT 190
QY 227 GCAGAGTGCTGCTCGTCACTCTTCCCAACAGCCTCATGTGCTGCCAGCTGCTCTGGC 286
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191 GCAGCAGTGATGCTCTCTCTGCTGTGAAGCCACACATCATCCCTTCATCTCTGA 250
QY 287 AGAGGTGGCTCTGCTCACCCTCAACACTCTGGTGTGGTGGCTGCTGGGCTGTC 346
|| || || || || || || || || || || || || || || || || || || || || ||
Db 251 TGAATAGGCGCGGACATTGAGGACAGACACATTTGGTGTCTCGCGCGCGCGGCTC 310
```

```
QY 347 TCTGAGCACCTTGAGGAGCTGCTG-----CCCCCAACACACAGGCTGCTGCGGGT 397
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 311 CATCAGCTCCATTGAGAAGAGCTGTTCAGCGTTTCGGCCAGCCCCCAGGCTCATCCCGCTG 370
QY 398 CTTGCCCAACCTGCCCTGTGTGTGCTCCAGGAAGGCCCATAGTATGCGGGCGGGCCGCCCA 457
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 371 CATGACCAACACTCCAGTCTGTGTGCGGGAGGGGGCCACCCTGTATGCCACAGCAGCA 430
QY 458 CGTGGGAGCAGCAGACCAAGCTTCGTCGACATCTGCTGGAGGCCCTGTGGCGGTGTGA 517
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 431 CGCCAGGTGAGGACGGGAGGCTCATGGACACCTGCTGAGCACGGTGGCTTCTGCAC 490
QY 518 GGAGTGTGCTGAAGCCTACGTCGACATCCACACTGGCTCAGTGGCAGTGGCGTGGCCTT 577
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 491 GGAGTGAAGAGGACCTGATTGATGCGCTCAGCGGGCTCAGTGGCAGCGGCCCGCCTA 550
QY 578 CGTGTGTGCAATTCCTCCAGGCCCCCTGGCTGAAGGAGCCGCTCAAGATGGGCATGCCACG 637
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 551 CGCATTCACAGCCCTGGATGCCCTGGCTGTGATGGGGGTGTGAAGATGGGACTTCCAAGCG 610
QY 638 CTGCGCCACCGCATCGCTGCCAGACCCCTGCTGGGAGCGGCCAAGATGCTGCTGCACGA 697
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 611 CTGGCAGTCCGCTCGGGGCCAGGCCCTCTCGGGGCTGCCAAGATGCTGCTGCACTC 670
QY 698 GGGCAACACACCCAGCCAGCTCGCTCAGACGTGTGCACCCCGGTGGCAGCACCATCTA 757
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 671 AGAACAGCACCAGGCCAGCTCAAGGACAACGTCAGCTCTCTGTTGGGGCCACCATCCA 730
QY 758 TGGACTCCACGCCCTGGAGCAGCGGGCTCGGAGCAGCCACCATGAGCGCGCTGGAGGC 817
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 731 TGCCTTGCAATGCTGGAGAGTGGGGCTTCGCCCTGCTCATCAACGCTGTGGAGGC 790
QY 818 TGCCACCTGCGGGCCAAAGGACT 841
Db 791 CTCCTGCATCCGACACGGGAGCT 814

RESULT 13
HUMP5CR
LOCUS Human pyrroline 5-carboxylate reductase mRNA, complete cds.
DEFINITION M77836
ACCESSION M77836.1 GI:189497
VERSION 1
KEYWORDS cytosolic enzyme; proline synthesis; pyrroline-5-carboxylate reductase.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1792)
AUTHORS Dougherty,K.M., Brandriss,M.C. and Valle,D.
TITLE Cloning human pyrroline-5-carboxylate reductase cDNA by complementation in Saccharomyces cerevisiae
JOURNAL J. Biol. Chem. 267 (2), 871-875 (1992)
MEDLINE 92112821
PUBMED 1730675
FEATURES
source Location/Qualifiers
1..1792
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HepG2"
/tissue_type="hepatoma"
12..971
/codon_start=1
/product="pyrroline-5-carboxylate reductase"
/protein_id="AAA36407.1"
/db_xref="GI:189498"
/translation="MSVGFIGAGQLAFALAKGFTAAGVLAHKIMASSPDMDLATVSA
LRKMGVLTTHNKETVQHSVLFVAVPHIIPFLDEIGADIEDRIHIVSCAAGVTTIS
SIEKLSLFRPAPRVIRCMWTNTPVVRREGATVATGTHAQVEDRLMEQLLSVGFCT
EVEEDLIDAVTGLSGPAYAFTALDALADGGVKMGLPRVLAVRGLAQLLGAALMLL
HSEHQPOLKDNVSPGATTHALHVLSESGGFRSLINAVEASCIRTRRELQSMADQEQ
VSPAARKTILDKVLDSPAGTALSPSGHTKLLPRSLAPAGK"
```


BASE COUNT	342 a	573 c	522 g	355 t
ORIGIN	Chromosome 17.			
Query Match	16.8%; Score 224.8; DB 9; Length 1792;			
Best Local Similarity	58.0%; Pred. No. 1.4e-30;			
Matches	466;	Conservative	0;	Mismatches 317; Indels 21; Gaps 3;
QY	56	GGCGTGGGCTTCGTGGGGCGGGCCGATGCGGGGGCCCATGCGGCGAGGCGCTCATCAG	115	
Db	14	GAGCGTGGGCTTCATCGCGCTGGCGCAGCTGGCTTTTGGCCGCAAGGGCTTCCACAGC	73	
QY	116	ACGAGGAAAGTGGAGCTCAGCACATAGTCCGAGTGCACCAACAGACAGAGACCATATG	175	
Db	74	ACAGCGCTTGGCTGCCCAAGATAAGTGTAGCTCCC---CAGACATGGACCTGGC	130	
QY	176	TCACCTTTCAAGTCTC-----TGGGTTGGCGGACCAACGACATCTCCACACAGGAGTGTCT	226	
Db	131	CACAGTTTCTGCTCTCAGGAAGATGGGGTGAAGTTGACACCCACACAGAGAGAGCGT	190	
QY	227	GCAGAGCTCCCTGCTGCTATCTTTGCCACAGAGCTCATGTGTCGCCAGTGTCTCGGCG	286	
Db	191	GAGGACAGATGATGTCTTCTTCTGCTGTGAAGCCACACATATCCCTTTTCATCCTGGA	250	
QY	287	AGAGGTGGCTCTGCTGGTCAACACTGAACATCTTGTGTCGCTGCTGGGTGCTC	346	
Db	251	TGAATAGCGCGGACATTTGAGGACAGACATTTGTGTGCTTGGCGGCGCGGCTGAC	310	
QY	347	TCAGAGACCTCTGGAGGAGCTGCTG-----CCCCCAACACACAGGCTGCTCGCGGT	397	
Db	311	CATCAGCTCCATTTGAGAAGAGCTGTGACGCTTTTGGGCGCACCCCGAGGGTCATCCGCTG	370	
QY	398	CTTGCCCAACCTGCGCTGTGGTCCAGAGGGGCCATAGTATGGCGGGCGGCGGCA	457	
Db	371	CATGACCAACACTCCAGTCGTGGTGGGAGGGGGGCCACCGTGTATGCCACAGGACGCA	430	
QY	458	CGTGGGAGCAGCAGACCAAGCTCTCTGACGATCTGTCGAGGCGCTGTGGGCGGTGGA	517	
Db	431	CGCCGAGGTGGAGGAGGGAGGCTCATGAGCAGCTGTGAGCAGGTTGGCTTCTGAC	490	
QY	518	GGAGTGGCTGAAGCCTACGTCGACATCCACACTGGGCTAGTGGGCTAGTGGCGTGGCCTT	577	
Db	491	GGAGTGAAGAGGACCTGATTGATGCGCTCACGGGCTCAGTGGCAGCGGCCCGCCTA	550	
QY	578	CGTGTGTCATCTCCGAGGCGCTGGTGAAGGAGCCCTCAAGATGGCATGGCCAGG	637	
Db	551	CCATCTCAGCGCTGGATGCCCTGGCTGATGGGGGTGTGAAGATGGGACTTCCAAAGCG	610	
QY	638	CCTGCGCCACCGCATCTGCTGCCAGACCTCTGCTGGGAGCGCCAGATGCTGTGCACGA	697	
Db	611	CCTGCGAGTCCGCTCGGGGCCAGGCGCTCTCTGGGGCTGCCAAGATGCTGTGCACTC	670	
QY	698	GGGCGCAACACCGCCAGCTGCGCTCAGACGTGTGACACCCCGGCTGGCACCACTCTA	757	
Db	671	AGAACAGCAGCCAGGCGCAGCTCAAGAGCAAGCTCAGCTCTCTGCTGGGGCCACCATCA	730	
QY	758	TGGACTCCACCGCTGGAGGCGGGCTCGGAGCAGCCACCATAGCGCGCTGGAGGC	817	
Db	731	TGCTGTGATGCTGGAGATGGGGGCTTCCGCTCCCTGCTCATCAACGCTGTGGAGGC	790	
QY	818	TGCCACCTCGCGGGGCCAAGGAGCT	841	
Db	791	CTCCTGTCATCCGACACGGGAGCT	814	
RESULT 14				
AF218000	Homo sapiens clone PP222 unknown mRNA.			
LOCUS	AF218000			
DEFINITION	Homo sapiens 1769 bp mRNA			
ACCESSION	AF218000			
VERSION	AF218000.1 GI:10441929			
KEYWORDS	FLI_CDNA			
SOURCE	Homo sapiens			
ORGANISM	Homo sapiens			

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1769)

Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P., Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y., Yu,J. and Han,L.H.

Novel Human cDNA clones with function of inhibiting cancer cell growth

Unpublished

2 (bases 1 to 1769)

Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P., Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y., Yu,J. and Han,L.H.

Direct Submission

Submitted (21-DEC-1999) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25 Ln 2200, Xie-tu Road, Shanghai 200032, People's Republic of China

Location/Qualifiers

1. 1769

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="pp222"

279. 1097

/codon_start=1

/product="unknown"

/protein_id="AAG17242.1"

/db_xref="GI:10441930"

/translation="MSVGFGLGQALAFALAKGFTAAAGVLAHKLIMASSPDMDLATVSA LRKMGKLVPHNKETVQHSDFVLAVKPHIIPILDEIGADIEDRHIVVSCAAGVTIS STEKILSAFPPAPRVTRCMTNPVVREGATVYATGTHAQVEDRGLMQLLSSVGEFT EVEDLIDAVTGLSGSPAYATFALDAGGVKMLGPLRLRALVRLGACALGACQDAA AURTAPRPAQGRQLSLWGHHPHLCAGBEMGLPLPAHORCGGLLHPHTGAAVHG"

BASE COUNT 374 a 533 c 530 g 332 t

ORIGIN

Query Match 16.0%; Score 213.8; DB 9; Length 1769;

Best Local Similarity 57.9%; Pred. No. 1.3e-28;

Matches 466; Conservative 0; Mismatches 317; Indels 22; Gaps 4;

QY 56 GC CGTGGGCTTCGTGGGGCGGGCGCATGCGGGGGCGCATCGCGGGCGGCTCATCAG 115

Db 281 GAGCGTGGGCTTCATCGCGCTGGCGCAGCTGGCTTTTGGCCGCAAGGGCTTCCACAG 340

QY 116 AGCAGGAAAGTGGAGCTCAGCACATCTGGCAGTGCACCAACAGACAGAGCACTATG 175

Db 341 AGCAGGCGTCTGGCTGCCACAAAGATAATAGTGTAGTCTCC---CAGACATGGAGCTGC 397

QY 176 TCACCTTTCAAGCTC-----TGGGTTGCCGACCGGCTCAACAGGAGTGTCT 226

Db 398 CACAGTTTCTGCTCTCAGGAAGATGGGGGTGAAGTTGACACCCCAACAGGAGCGGT 457

QY 227 GCAGAGCTGCCTGCTCATCTTTGCCACCAAGCTCATGTGCTGCCAGCTGTCTCTGCG 286

Db 458 GCAGCAGTGTGCTCTCTCTGCTGTGAAGCCACACATCATCCCTTCTCATCTGGA 517

QY 287 AGAGGTGGCTCTGCTGGTCCACACTCAACATCTTGTGTCGCTGGCTGTGGGGTGTG 346

Db 518 TGAATAGGCGGCGACATTTGAGCAGACACATTTGTTGTTGCTGCTGCGCGCGGGGTGAC 577

QY 347 TCTGAGCACCTGGAGGAGCTGTG-----CCCCCAACACAGCGGCTGTGCGGT 397

Db 578 CATCAGCTCCATTTGAGAAGAGCTGTACAGGTTTGGCCAGCCCGGAGGCTCATCCGTG 637

QY 398 CTTGCCCAACCTGCCCTGTGTGTCAGGAAGGGCCATAGTATGCGCGGGCGGCGCA 457

Db 638 CATGACCAACACTCCAGTGTGTCGGGAGGGGCCACCGTGTATGCCACAGGACGCA 697

QY 458 CGTGGGAGCAGGACCAACAGCTCTCTGACGATCTGTGTGGAGGCTGTGGCGGTGTGA 517

Db 698 CCGCCAGGTGGAGGAGGAGGCTCATGAGAGAGCTGTGAGCAGTGGGCTTCTGAC 757

QY 518 GGAGGTGCCTGAAGCCTACGTCACATCCACACTGGCCTCAGTGGCGTGGCGTGCCTT 577

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:52:06 ; Search time 26 Seconds
(without alignments)
310.072 Million cell updates/sec

Title: US-09-806-536A-14
Perfect score: 1386
Sequence: 1 MAAAEPSRRVGFVGAGRMA.....AATMSAVEAATCRAKELSRK 274

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	560.5	40.4	315	3	US-09-099-676-3
2	560.5	40.4	315	4	US-09-565-910-3
3	537.5	38.8	314	3	US-09-099-676-1
4	537.5	38.8	314	4	US-09-565-910-1
5	297.5	21.5	304	1	US-08-665-716-2
6	269	19.4	282	4	US-09-134-001C-3155
7	116	8.4	144	4	US-08-936-165A-445
8	91.5	6.6	436	3	US-08-669-378-4
9	88	6.3	342	4	US-09-364-230-6
10	87.5	6.3	565	3	US-09-008-481A-6
11	87.5	6.3	565	4	US-09-195-666A-5
12	87.5	6.3	565	4	US-09-195-666A-49
13	87.5	6.3	565	4	US-09-309-592-6
14	87.5	6.3	565	4	US-09-635-705-5
15	87.5	6.3	565	4	US-09-635-705-49
16	87.5	6.3	565	4	US-09-634-858A-5
17	87.5	6.3	565	4	US-09-634-858A-49
18	87.5	6.3	1024	4	US-09-562-737-44
19	87	6.3	528	4	US-08-928-213B-8
20	87	6.3	949	4	US-09-196-387-10
21	87	6.3	1327	4	US-09-196-387-2
22	86	6.2	551	3	US-08-886-886-15
23	85.5	6.2	746	2	US-08-785-431-4
24	85.5	6.2	746	4	US-09-205-048-4
25	85.5	6.2	788	2	US-08-785-431-2
26	85.5	6.2	788	4	US-09-205-048-2
27	85	6.1	3729	2	US-08-804-227C-4

28	84.5	6.1	436	3	US-08-669-378-6	Sequence 6, Appli
29	84.5	6.1	436	3	US-08-669-378-10	Sequence 10, Appl
30	84.5	6.1	638	4	US-09-413-814-102	Sequence 102, App
31	83	6.0	2475	4	US-09-413-814-48	Sequence 48, Appl
32	82.5	6.0	333	4	US-09-338-420-4	Sequence 4, Appli
33	82.5	6.0	386	4	US-09-338-420-2	Sequence 2, Appli
34	82.5	6.0	573	3	US-09-008-481A-7	Sequence 7, Appli
35	82.5	6.0	573	4	US-09-195-666A-50	Sequence 50, Appl
36	82.5	6.0	573	4	US-09-309-592-7	Sequence 7, Appli
37	82.5	6.0	573	4	US-09-635-705-50	Sequence 50, Appl
38	82.5	6.0	573	4	US-09-634-858A-50	Sequence 50, Appl
39	82.5	6.0	741	2	US-08-849-480A-9	Sequence 9, Appli
40	82	5.9	5405	4	US-08-718-388-9	Sequence 9, Appli
41	81.5	5.9	436	3	US-08-669-378-2	Sequence 2, Appli
42	81.5	5.9	436	3	US-08-669-378-8	Sequence 8, Appli
43	81	5.8	345	4	US-09-364-230-8	Sequence 8, Appli
44	80	5.8	382	4	US-09-586-719-8	Sequence 8, Appli
45	79.5	5.7	781	4	US-09-766-387-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-099-676-3
; Sequence 3, Application US/09099676
; Patent No. 6100075
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
; TITLE OF INVENTION: HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/099,676
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carrione, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0532 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-855-0572
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 315 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 189498

Query Match 40.4%; Score 560.5; DB 3; Length 315;

[illegible]

RESULT 2

```

US-09-565-910-3
: Sequence 3, Application US/09565910
: Patent No. 6268192
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Corley, Neil C.
: APPLICANT: Raughn, Mariah R.
: TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
: NUMBER OF SEQUENCES: 3
: TITLE OF INVENTION: HOMOLOG
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: Windows
: SOFTWARE: FastSeq for Windows Version 2.0b
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/565,910
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/099,676
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Ceirone, Michael C
: REGISTRATION NUMBER: 39,132
: REFERENCE/DOCKET NUMBER: PF-0532 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-855-0572
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 315 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: GENBANK
: CLONE: 189498
: US-09-565-910-3

```

Query Match	40.4%;	Score 560.5;	DB 4;	Length 315;
Best Local Similarity	46.1%;	Pred. 0.1.3e-53;		
Matches 123;	Conservative 46;	Mismatches 87;	Indels 11;	Gaps 5;
QY	11	VGFVCGAGMAGAAGLLTRACKVKAQHILASPTDRNLCHFQAL---	GCRTTHSHNQEVQLQ	67
DB	3	VGFICAGQL--AFAGFTTAAGVLAHKKMASSP-DMDLATVSALRKMGVKLLPHNKETVQ		59
QY	68	SCLLVIFATKPHVPAVLAIEVAPVVTTEHILVSVAAGVSLSTLEELP---	PNTRLRVLRLV	124
DB	60	HSDLVFLAVKPHIIPFLDEIGADIEDRHVWVSCAAGVTISSTIEKLSAFRRPAPRVIRCM		119
QY	125	PNLPCVWOBGATVMARGHVSSSTKLLQHLLEACGRCEVPAYVDIHGTLSSGSGVAFV		184
DB	120	TNTPVVVREGATVYATGTHAQVDEGRLMEQLLSTVVGCTEVEDLLDVTGLSSGSPAY-		178
QY	185	CAFSALAEAGVAKMGMPSSLAHRTAAOTLLGTAKMLLHBEGQHPAQRLSDVCTPFGGTTYG		244
DB	179	-AFTALDADGGVKGGLPRRLAVRIGAAQLLGAAKMILLHSEHQPGQIKDNVSSPGGATTHA		237
QY	245	LHAEQGGDLRAATMSAVEAATCRAKEL		271
DB	238	LHVLESGGFRSLLINAVEASCIRREL		264

RESULT 3

```

US-09-099-676-1
:
: Sequence 1, Application US/09099676
: Patent No. 6100075
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Corley, Neil C.
: APPLICANT: Baughn, Maria R.
: TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
: TITLE OF INVENTION: HOMOLOG
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows
: SOFTWARE: FastSQ for Windows Version 2.0b
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/099,676
: FILING DATE: HEREWITH
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Cerrone, Michael C
: REGISTRATION NUMBER: 39,132
: REFERENCE/DOCKET NUMBER: PF-0532 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-855-0572
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 314 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: PROSNON01
: CLONE: 2278458
: IS-09-099-676-1

```


QY 3 AAEPSRRVGVFAGRMAGAIAOGLIRA-GKVEAQHILASAPTRNLCHE--QALGORTT 59
Db 5 AKEGSELTAIVGCTGTMGIAILSLGILASLDEIHAFN-SQSETDETPTSKLPTKFIACVRS 63
QY 60 HSNQEVLSQCL-----LVIFATKPHVLPVLAIEVAVPVVTE-H 96
Db 64 PKGAEKIKKALSPYKTPVKIIQSDNVNACREADVVLGCRPYMAEGLIBEGWVDALUGK 123
QY 97 ILVSVAAVGSVSTLEELL-----PPN-----TRVLRVLPNLCVQVQGAIVMARGR-HVG 145
Db 124 LLISILAGVPAEQIYGYMGKTPVNPKEGLCVVRAMPNTASGIRESMVIATSSPPLS 183
QY 146 SSETKLLQHLLEACGRCEVEPEAYVDHTGLSGGVAFVCAFSALAGAVKMGMPSSLA 205
Db 184 ATTSLITWIFKRIGDVVQLPAATMDASTALCGSGPAFFALILEAAIDGAVAMGLPRAEA 243
QY 206 HRIAQAQTLTGAKMLHEGQHPAQLRSDVCTPGGTTIYGLHALEOGLRAATSAVBAAT 265
Db 244 QRMAAQTMKGAAGLVL-SGEHPALLKDKVTTPGCTTIGGLMVLVEGGVGTVARAVREAT 302

RESULT 6

US-09-134-001C-3155
; Sequence 3155, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3155
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3155

Query Match 19.4%; Score 269; DB 4; Length 282;
Best Local Similarity 26.9%; Pred. No. 1.7e-21;
Matches 71; Conservative 54; Mismatches 137; Indels 2; Gaps 2;

QY 13 FVAGRMAGAIAOGLIRAGKVEAQHI-LASAPTRNLCHE--QALGORTTTHSNQEVLSQCL 70
Db 16 FYGAGNMAQAIFGTGIINSNNLNDIYLTNKSNEQALKSFAEKLGVNYSYDDDEALLKDDAD 75
QY 71 LVIFATKPHVLPVLAIEVAVPVVTEHILVSVAAVGSVSTLEELLPPNTRVLRVLPNLCV 130
Db 76 YVFLGTPKPHDFENLARIREHTINDRFTISIMAGLSIDYIROLNTNPNLARIMPNTNAQ 135
QY 131 VQEGAIVMARGRIVGSSETKLLQHLLEACGRCEVEPEAYVDHTGLSGGVAFVCAFSFA 190
Db 136 VGHSVTGISFNNFDPKSNVEDELINAFSGVIEVEEHLHQVATLGGSGAFLYHVFEQ 195
QY 191 LAEGAVKMGMPSSLAHRIAQAQTLTGAKMLHEGQHPAQLRSDVCTPGGTTIYGLHALEQ 250
Db 196 YVRAGTELGLERNOVEESIRNLIGTSKMIERSDLSMSQLRKNITSKGGTTQAGLDLSQ 255
QY 251 GGLRAATMSAVEATCRAKELSRK 274
Db 256 YDIVSMFECLGAAVNRSMELSHK 279

RESULT 7

US-08-936-165A-445
; Sequence 445, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:

; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Richard
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,
; TITLE OF INVENTION: Polypeptides and Their Uses
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,165A
; FILING DATE: 24-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027,032
; FILING DATE: 24-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 445:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; US-08-936-165A-445

Query Match 8.4%; Score 116; DB 4; Length 144;
Best Local Similarity 26.8%; Pred. No. 4.7e-05;
Matches 30; Conservative 24; Mismatches 56; Indels 2; Gaps 2;

QY 16 AGRMAGATAOGLIRAGKVEAQHI-LASAPTRNLCHE--QALGORTTHSNQEVLSQCLVI 73
Db 12 AGNMAQAIFGTGIINSNNLNDIYLTNKSNEQALKSFAEKLGVNYSYDDDEALLKDDADYF 71
QY 74 FATKPHVLPVLAIEVAVPVVTEHILVSVAAVGSVSTLEELLPPNTRVLRVLP 125
Db 72 LGTRKPHDFDALATRIKPHITKDXCFMSIMAGIPTDYIXQOLECONPARIMP 123

RESULT 8

US-08-669-378-4
; Sequence 4, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:

; APPLICANT: Moeckel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
; DEHYDRATASE

NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/08/669,378
FILING DATE: 20-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DE95/00017
FILING DATE: 09-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 44 00 926.7
FILING DATE: 14-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 016881/0142
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 9041136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 436 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-669-378-4

Query Match 6.6%; Score 91.5; DB 3; Length 436;
Best Local Similarity 20.3%; Pred. No. 0.13; Mismatches 133; Indels 95; Gaps 18;
Matches 71; Conservative 50;
QY 3 AAEPSPRR-VGFYAGRMAGAIQAOLIRA-----GKVEAQHILASAPTDRLNLC---F 51
DB 79 AQSPQEQRDAGIVAAS-AGNHAQGVAYVCKSLGVQGIYVVPQTPKQRDRIMVHGGEF 136
QY 52 QAL-----GCRTHSNQEVLSQCLLVIFATKPHVL--PAVLAEVAPVVT-----E 95
DB 137 VSLVVTGNNFDEASAAHEDAERTGATLIEPFDARNTVIGCGTAAEILSQTSMGKSAD 196
QY 96 HILVSAAGVSLSTLEEL---LPNTRVLRVLP-----NLPCVVOEGAIYV----- 138
DB 197 HVMVPVGGGLLAGVSVYADMAPRTAIVIEPAGAAASMOAALHNGGPITLETVDPDFVG 256
QY 139 ARGHRVGSSTKLQ-----HLEAC--GRCEVPEAYVDIHTGLSGSVAFVCAFSF- 189
DB 257 AEVKRVGDLNTYIVKNGRVHMSATEGAVCTEMLDLYONEGIIAEPAGALSIAGLKEM 316
QY 190 ALAEGAVKM-----GMPSSLAHRIAQTLLGTAKMLLHG-----QHPAQLR--- 231
DB 317 SFAPGSVVVCIISGGNNDLRYAEIAE-----RSLVHRGLKHVFLVNFPPQKPGQLRHFL 370
QY 232 SDVCTPG-----GTTIYGLHLEOQGLRAATMSAVEAA 264
DB 371 EDILPGDDITLFLKLRNNRETGTALVGIHLEASGLDLSLERMEESA 419

RESULT 9
US-09-364-230-6
; Sequence 6, Application US/09364230
; Patent No. 6348339
; GENERAL INFORMATION:

; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: Enzymes Involved in Degradation of Branched-Chain Amino Acids
; FILE REFERENCE: BB-1178
; CURRENT APPLICATION NUMBER: US/09/364,230
; CURRENT FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,990
; EARLIER FILING DATE: July 31, 1998
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Oryza sativa
; US-09-364-230-6
Query Match 6.3%; Score 88; DB 4; Length 342;
Best Local Similarity 25.4%; Pred. No. 0.21;
Matches 60; Conservative 36; Mismatches 90; Indels 50; Gaps 13;
QY 2 AAAEPSRRVGVGAGRMAGAIQAOLIRAG-KVEAQHILASAPTDRLNLCFQALGCRTH 60
DB 26 AAVQSQLENVGFGLGNMGAHMARLVNMGYKVTVDV-----NENTMKKFSDDGIPTKL 80
QY 61 SNOEVLQSCLLVI--FATKPHVLP-----AVLAE---VAPVVTTEHILV-----SVAA 103
DB 81 SPLEVSKSDVVITMLPSSAHVLDVYNGRNLGLGCLGPWLYIDSTVDPQTSRKISM 140
QY 104 GVSLSLEELPPNTRVLRVLPNLCVV---QEGAIVMARGHRHVGSSSTKLLQH-LLBAC 159
DB 141 DISRCTLEKKPYAEKPMMLDAPVPGVPPAEAGKLTFLVG---GSEAYLAANPLLSM 197
QY 160 GRCEVPEAYVDIHTGLSGS-VAPVC-----AFSALAEAGVAKMGPSSL 204
DB 198 GK-----RTIYCGGAGNGSVAKICNNMAMGISMVGLVSEAFALGQ-NLGIKASV 244

RESULT 10
US-09-008-481A-6
; Sequence 6, Application US/09008481A
; Patent No. 6087151
; GENERAL INFORMATION:
; APPLICANT: Ario, Takeshi
; APPLICANT: TANIAL, Madoka
; APPLICANT: TORIGOE, Kakuji
; APPLICANT: KURIMOTO, Masashi
; TITLE OF INVENTION: DNA CODING FOR MAMMALIAN L-
; TITLE OF INVENTION: ASPARAGINASE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,481A
; FILING DATE: 16-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/598,369
; FILING DATE: 08-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 95-42564
; FILING DATE: 08-FEB-1995
; ATTORNEY/AGENT INFORMATION:

NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,981
REFERENCE/DOCKET NUMBER: ARI0-1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-008-481A-6

Query Match 6.3%; Score 87.5; DB 3; Length 565;

Best Local Similarity 24.5%; Pred. No. 0.52;

Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;

QY 25 QGLIRAGKVEAQHILASAPTRNLCHFQALGCRTHSHNQEVLOS-CLLVIFATKPHVLP 83

DB 155 ENLLGALLVAGQYIIPEVCLFMNSQLFR--GNRVTKVDSQKFEAFC-----SPNLS- 204

QY 84 VLAEPVPTTEHILVSVAAGVSLSTLEELLPPNTRVRLVLPNLCV-----QEGA 135

DB 205 -LATVGADVTIADLVKRVNWKDPLVHSHNMEHDVALLRLYPCIPASLVRAFLQPLKGV 263

QY 136 IVMARGRHVSSETKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSV--A 182

DB 264 VLETFGSGNGPSKPDLLQELRAAQRGLIWNCSQCLRGSVTPGYA---TSLAGANIVSG 320

QY 183 FVCAFEALAGAVKMGMS--SLAHRTAAQTLLGTAKMLLHGOHPAQLRSDVCTPGGTT 241

DB 321 LDMTSEAAKLVSVLGPELSLERR---QELL--AKDLRGEMTLPTADLHQSSPPGSTL 375

QY 242 IYGLHAL-----EOGGLRAATMSAVEAATCRAKEL 271

DB 376 GQGVARLFSLFGCQEDSVQDVMPSLALAHAGEL 412

RESULT 11

US-09-195-666A-5

Sequence 5, Application US/09195666A

Patent No. 6140101

GENERAL INFORMATION:

APPLICANT: ARI0, Takeshi

APPLICANT: TANAI, Modaka

APPLICANT: YAMAMOTO, Kozo

APPLICANT: KURIMOTO, Masashi

TITLE OF INVENTION: POLYPEPTIDES HAVING L-ASPARAGINASE ACTIVITY

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.

STREET: 624 Ninth Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/195,666A

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/869,927

FILING DATE: 05-JUN-1997

APPLICATION NUMBER: JP 168,172/96

FILING DATE: 07-JUN-1996

ATTORNEY/AGENT INFORMATION:

NAME: YUN, Allen C.

REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: ARI0-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-195-666A-5

Query Match

Best Local Similarity 24.5%; Score 87.5; DB 4; Length 565;

Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;

QY 25 QGLIRAGKVEAQHILASAPTRNLCHFQALGCRTHSHNQEVLOS-CLLVIFATKPHVLP 83

DB 155 ENLLGALLVAGQYIIPEVCLFMNSQLFR--GNRVTKVDSQKFEAFC-----SPNLS- 204

QY 84 VLAEPVPTTEHILVSVAAGVSLSTLEELLPPNTRVRLVLPNLCV-----QEGA 135

DB 205 -LATVGADVTIADLVKRVNWKDPLVHSHNMEHDVALLRLYPCIPASLVRAFLQPLKGV 263

QY 136 IVMARGRHVSSETKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSV--A 182

DB 264 VLETFGSGNGPSKPDLLQELRAAQRGLIWNCSQCLRGSVTPGYA---TSLAGANIVSG 320

QY 183 FVCAFEALAGAVKMGMS--SLAHRTAAQTLLGTAKMLLHGOHPAQLRSDVCTPGGTT 241

DB 321 LDMTSEAAKLVSVLGPELSLERR---QELL--AKDLRGEMTLPTADLHQSSPPGSTL 375

QY 242 IYGLHAL-----EOGGLRAATMSAVEAATCRAKEL 271

DB 376 GQGVARLFSLFGCQEDSVQDVMPSLALAHAGEL 412

RESULT 12

US-09-195-666A-49

Sequence 49, Application US/09195666A

Patent No. 6140101

GENERAL INFORMATION:

APPLICANT: ARI0, Takeshi

APPLICANT: TANAI, Modaka

APPLICANT: YAMAMOTO, Kozo

APPLICANT: KURIMOTO, Masashi

TITLE OF INVENTION: POLYPEPTIDES HAVING L-ASPARAGINASE ACTIVITY

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.

STREET: 624 Ninth Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/195,666A

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/869,927

FILING DATE: 05-JUN-1997

APPLICATION NUMBER: JP 168,172/96

FILING DATE: 07-JUN-1996

ATTORNEY/AGENT INFORMATION:

NAME: YUN, Allen C.

REGISTRATION NUMBER: 37,971

REFERENCE/DOCKET NUMBER: ARI0-2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)628-5197
TELEFAX: (202)737-3528
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-195-666A-49

Query Match 6.3%; Score 87.5; DB 4; Length 565;
Best Local Similarity 24.5%; Pred. No. 0.52;
Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;
QY 25 OGLIRAGKVEAQHILASATDRNLCHFOALGCRTHSHNOEVLQS-CLLVIFATKPHVLP 83
Db 155 ENLLGALLVAGQYIIPEVCLFMNSQLFR--GNRVTKVDSQKFEAFC-----SPNLS- 204
QY 84 VLAEPVPTTEHILSVAAAGVSLSTLELLPNTRLVRLVLPNLPV-----QEGA 135
Db 205 -LATVGADVTIAMDILVRKVNWKDPLVHNSMEHDVALLRLPGIPASIVRAFLQPLKGV 263
QY 136 IVMARGRHVGSSETKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSV--A 182
Db 264 VLETFGSGNGSPKPDLLQELRAAQAORGLIMVNCSQLRGSVTPGYA---TSLAGANIVSG 320
QY 183 FVCAFSEALAEAGVAKMGMP--SLAHRIAQTLLGTAKMLLHGQHPAQLRSDVCTPGGTT 241
Db 321 LDMTSEAALAKLSYVLGLPELSLERR---QELL--AKDLRGEMTLPTADLHQSPPGSL 375
QY 242 IYGLHAL-----EOGGLRAATMSAVEAATCRAKEL 271
Db 376 GOGVARLFSFGCEEDSVQDAVMPSLALAHAGEL 412

RESULT 13
US-09-309-592-6
Sequence 6, Application US/09309592
Patent No. 6274367
GENERAL INFORMATION:
APPLICANT: ARIO, Takeshi
APPLICANT: TANIAI, Madoka
APPLICANT: TORIGOE, Kakuji
APPLICANT: KURIMOTO, Masashi
TITLE OF INVENTION: DNA CODING FOR MAMMALIAN L-
TITLE OF INVENTION: ASPARAGINASE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/309,592
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/008,481
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 95-42564
FILING DATE: 08-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,981

REFERENCE/DOCKET NUMBER: ARIO-1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-309-592-6

Query Match 6.3%; Score 87.5; DB 4; Length 565;
Best Local Similarity 24.5%; Pred. No. 0.52;
Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;
QY 25 OGLIRAGKVEAQHILASATDRNLCHFOALGCRTHSHNOEVLQS-CLLVIFATKPHVLP 83
Db 155 ENLLGALLVAGQYIIPEVCLFMNSQLFR--GNRVTKVDSQKFEAFC-----SPNLS- 204
QY 84 VLAEPVPTTEHILSVAAAGVSLSTLELLPNTRLVRLVLPNLPV-----QEGA 135
Db 205 -LATVGADVTIAMDILVRKVNWKDPLVHNSMEHDVALLRLPGIPASIVRAFLQPLKGV 263
QY 136 IVMARGRHVGSSETKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSV--A 182
Db 264 VLETFGSGNGSPKPDLLQELRAAQAORGLIMVNCSQLRGSVTPGYA---TSLAGANIVSG 320
QY 183 FVCAFSEALAEAGVAKMGMP--SLAHRIAQTLLGTAKMLLHGQHPAQLRSDVCTPGGTT 241
Db 321 LDMTSEAALAKLSYVLGLPELSLERR---QELL--AKDLRGEMTLPTADLHQSPPGSL 375
QY 242 IYGLHAL-----EOGGLRAATMSAVEAATCRAKEL 271
Db 376 GOGVARLFSFGCEEDSVQDAVMPSLALAHAGEL 412

RESULT 14
US-09-635-705-5
Sequence 5, Application US/09635705
Patent No. 6368845
GENERAL INFORMATION:
APPLICANT: ARIO, Takeshi
APPLICANT: TANIAI, Madoka
APPLICANT: YAMAMOTO, Kozo
APPLICANT: KURIMOTO, Masashi
TITLE OF INVENTION: POLYPEPTIDES HAVING L-ASPARAGINASE ACTIVITY
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 624 Ninth Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/635,705
FILING DATE: 09-Aug-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/195,666
FILING DATE: <Unknown>
APPLICATION NUMBER: JP 168,172/96
FILING DATE: 07-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: ARIO=2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)628-5197
TELEFAX: (202)737-3528
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-635-705-5

Query Match 6.3%; Score 87.5; DB 4; Length 565;
Best Local Similarity 24.5%; Pred. No. 0.52;
Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;
QY 25 QGLIRAGKVEAQHILASAPDRNLCHFQALGCRTHSNQEVLOS-CLLVIFATKPHVLP 83
Db 155 ENLIGALLVAGQYIPEVCLFMNSQLFR--GNRVTKVDSOKFEAFC-----SPNLS- 204
QY 84 VLAEPVAVVTTTEHILSVSAAGVSLSTLEELLPPNTRVLRVLPNLPV-----QEGA 135
Db 205 -LATVGADVTIANDLVKRVNWKDPLVVHNSMEHDVALLRYPASLVRALQPLKGV 263
QY 136 IVMARGHVSSETKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSV--A 182
Db 264 VLETFGSGNGPSKPDLLQELRAAQRGLIMVNCSQLRGSVTPGYA---TSLAGANIVSG 320
QY 183 FVCAFSEALAEAGVAKMGMP-SLAHRIAQTLLGTAKMLLHGOHPAQLRSDVCTPGGTT 241
Db 321 LDMTSEALAKLSVGLPELSLERR--QELL--AKDLRGEMTLPTADLHQSPPGSTL 375
QY 242 IYGLHAL-----EOGGLRAATMSAVEAATCRAKEL 271
Db 376 GOGVARLFLSGCOEDSVQDVMPSLALALAHAGEL 412

RESULT 15

US-09-635-705-49

Sequence 49, Application US/09635705
Patent No. 6368845

GENERAL INFORMATION:

APPLICANT: ARIO, Takeshi

TANAI, Modaka

YAMAMOTO, Kozo

KURIMOTO, Masashi

TITLE OF INVENTION: POLYPEPTIDES HAVING L-ASPARAGINASE ACTIVITY

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.

STREET: 624 Ninth Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/635,705

FILING DATE: 09-Aug-2000

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/195,666

FILING DATE: <Unknown>

APPLICATION NUMBER: JP 168,172/96

FILING DATE: 07-JUN-1996

ATTORNEY/AGENT INFORMATION:

NAME: YUN, Allen C.

REGISTRATION NUMBER: 37,971

REFERENCE/DOCKET NUMBER: ARIO=2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)628-5197
TELEFAX: (202)737-3528
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 49:
US-09-635-705-49

Query Match 6.3%; Score 87.5; DB 4; Length 565;
Best Local Similarity 24.5%; Pred. No. 0.52;
Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;
QY 25 QGLIRAGKVEAQHILASAPDRNLCHFQALGCRTHSNQEVLOS-CLLVIFATKPHVLP 83
Db 155 ENLIGALLVAGQYIPEVCLFMNSQLFR--GNRVTKVDSOKFEAFC-----SPNLS- 204
QY 84 VLAEPVAVVTTTEHILSVSAAGVSLSTLEELLPPNTRVLRVLPNLPV-----QEGA 135
Db 205 -LATVGADVTIANDLVKRVNWKDPLVVHNSMEHDVALLRYPASLVRALQPLKGV 263
QY 136 IVMARGHVSSETKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSV--A 182
Db 264 VLETFGSGNGPSKPDLLQELRAAQRGLIMVNCSQLRGSVTPGYA---TSLAGANIVSG 320
QY 183 FVCAFSEALAEAGVAKMGMP-SLAHRIAQTLLGTAKMLLHGOHPAQLRSDVCTPGGTT 241
Db 321 LDMTSEALAKLSVGLPELSLERR--QELL--AKDLRGEMTLPTADLHQSPPGSTL 375
QY 242 IYGLHAL-----EOGGLRAATMSAVEAATCRAKEL 271
Db 376 GOGVARLFLSGCOEDSVQDVMPSLALALAHAGEL 412

Search completed: July 21, 2003, 09:56:32

Job time : 28 secs

17	115.6	8.6	482	12	US-09-918-995-31035	Sequence 31035, A
18	84.4	6.3	810	11	US-09-738-626-460	Sequence 460, Ap
19	84.4	6.3	3309400	11	US-09-738-626-1	Sequence 1, Appl
20	83.2	6.2	415	12	US-09-918-995-17742	Sequence 17742, A
21	73.4	5.5	888	11	US-09-726-397A-4	Sequence 4, Appl
22	73.4	5.5	888	11	US-09-712-363-28	Sequence 28, Appl
23	73.4	5.5	86114	14	US-10-080-170-648	Sequence 648, App
24	72.8	5.4	931	15	US-10-027-632-161469	Sequence 161469, Ap
25	72.8	5.4	931	15	US-10-027-632-161470	Sequence 161470, Ap
26	70.6	5.3	780	11	US-09-974-300-343	Sequence 343, App
27	69.5	5.2	843	15	US-10-037-632-170626	Sequence 170626, Ap
28	62	4.6	855	15	US-10-128-714-2252	Sequence 2252, Ap
29	62	4.6	855	15	US-10-128-714-7252	Sequence 7252, Ap
30	59.8	4.5	1014	11	US-09-070-927A-157	Sequence 157, App
31	54.6	4.1	2908	15	US-10-128-714-252	Sequence 252, App
32	54.6	4.1	2908	15	US-10-128-714-5252	Sequence 5252, Ap
33	54.4	4.1	908	15	US-10-128-714-1252	Sequence 1252, Ap
34	54.4	4.1	908	15	US-10-128-714-6252	Sequence 6252, Ap
35	53.6	4.0	2209	11	US-09-726-397A-2	Sequence 2, Appl
36	52.2	3.9	334	11	US-09-796-632-6261	Sequence 6261, Ap
37	52.2	3.9	334	15	US-10-040-862-6261	Sequence 6261, Ap
38	48.2	3.6	2817	15	US-10-156-761-390	Sequence 390, App
39	47.4	3.5	588	15	US-10-156-761-1531	Sequence 1531, Ap
40	47.4	3.5	9025608	15	US-10-156-761-1	Sequence 1, Appl
41	47.2	3.5	795	15	US-10-156-761-1729	Sequence 1729, Ap
42	46.4	3.5	1359	15	US-10-156-761-6088	Sequence 6088, Ap
43	45.6	3.4	834	15	US-10-156-761-1113	Sequence 1113, Ap
44	44.4	3.3	903	15	US-10-156-761-2474	Sequence 2474, Ap
45	44	3.3	1645	15	US-10-037-270-807	Sequence 807, App


```
Db 791 CTCCTGCATCCGACACGGGAGCT 814
RESULT 6
US-10-161-418A-1
; Sequence 1, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: PSYCRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1792
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-161-418A-1
Query Match 16.8%; Score 224.8; DB 15; Length 1792;
Best Local Similarity 58.0%; Pred. No. 1.6e-45;
Matches 466; Conservative 0; Mismatches 317; Indels 21; Gaps 3;
QY 56 GCGCGTGGCTTCGTGGCGGGGCGCATGGCGGGGCGCATGGCGAGGCGCTCATCAG 115
Db 14 GAGCGTGGCTTCATCGCGGCTGGCGAGCTGGCTTTCCCTTGGCCAAAGGCTTCACAGC 73
QY 116 AGCAGAGAAAGTGAAGTTCAGACATATGCGCCAGTGCACCAACAGACAGAGACCTATG 175
Db 74 AGCAGGCGTCTGGCTGCCACAGATATGCTAGCTCC--CAGACATGGACCTGGC 130
QY 176 TCACATTCAGCTC-----TGGGTTGCCGACGAGCTCCACACAGGAGGCTCT 226
Db 131 CACAGTTTCTGCTCAGGAAGATGGGGGTGAAGTTGACACCCCAACAAAGGAGCGGT 190
QY 227 GCAGAGCTGCTGCTGCTCATCTTTGCCACCAAGCCTCATGTGCTGCCAGCTGTCTGGC 286
Db 191 GCAGCACAGTATGCTCTCTCTGCTGTGAAGCCACACATCATCTCCCTTCATCTCGGA 250
QY 287 AGAGTGGCTCTGTGGTCCACCATGACACATCTTGGTGTCCGTGGCTGTCTGGGGTGT 346
Db 251 TGAATAGGCGCGGACATTTGAGGACAGACACATTTGCTGTCTGCGGCGCGGCGTCCAC 310
QY 347 TCTGAGCACCTGGAGGAGCTGCTG-----CCCCCAACACACAGGCTGCTGCGGGT 397
Db 311 CATCAGCTCCATTGAGAGAACTGTACGCTTTCGGCCAGCCCCCAGGGTCTCCGCTG 370
QY 398 CTTGCCCAACCTGCTGTGTGTCAGGAGGCGCCATAGTGTATGGCGGGGCGGCCCA 457
Db 371 CATGACCACTCCAGTCTGTGTCGGGAGGCGCCACCGTGTATGCCAGGCAACGCA 430
QY 458 CGTGGGACACGAGACCAACTCTGCAGCATCTGTGAGGCGCTGTGGCGGTGTGA 517
Db 431 CGCCCCAGGTGGAGAGCGGAGGCTCATGGAGCAGCTGTGAGCAGGTTGGGCTTCTGCAC 490
QY 518 GGAGGTGCTTAAGCCTACGCTGCATCCACATCGCCTCAGTGGCAGTGGCGTGGCGCTT 577
Db 491 GGAGGTGGAAGAGGACCTGATTGATGCGCTCAGGGGCTCAGTGGCAGGCGCGCGCTA 550
QY 578 CTTGTGTGATTTCTCGAGGCGCTTGGCTGAAGAGCGCGTCAAGATGGGCATGCCAGCAG 637
Db 551 CGCATTCACAGCGCTGATGCTGCTGTGCTGATGGGCTGTGAAGATGGGACTTCCAAAGGC 610
QY 638 CTTGGCCCAACCGCATCGCTGCCAGACCTGCTGGGAGCGGCAAGATGCTGCTGCACGA 697
Db 611 CTTGGCAGTCCGCTCGGGGCGCCAGGCCCTCTCTGGGGGCTGCCAAGATGCTGCTGCATC 670
QY 698 GGGCCAAACCCAGCCAGCTGCGCTCAGACGTTGTACCCGGGTGGCACCACTCTA 757
Db 671 AGAACAGCACCCAGGCGGCTCAAGGACAAACGTCAGCTCTCTCTGGTGGGCGCCACCA 730
QY 758 TGGACTCCAGCGCCTGAGCAGGCGGCTGCGAGCAGCCACCATGAGGCGCGTGGAGGC 817
Db 731 TGCCTTGCATGTCTGAGAGTGGGGCTTCCGCTCCCTGCTCATCAAGCTGTGGAGGC 790
QY 818 TGGCACTTGGCGGCGCAAGGAGCT 841
Db 791 CTCCTGCATCCGACACGGGAGCT 814
RESULT 7
US-10-161-418A-5
; Sequence 5, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: PSYCRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1715
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-161-418A-5
Query Match 16.0%; Score 214.4; DB 15; Length 1715;
Best Local Similarity 58.0%; Pred. No. 5.4e-43;
Matches 466; Conservative 0; Mismatches 316; Indels 22; Gaps 4;
QY 56 GCGCGTGGCTTCTGTGGCGGGGCGCATGGCGGGGCGCATGGCGAGGCGCTCATCAG 115
Db 281 GAGCGTGGCTTCTATCGCGCTGGCCAGCTGGCTTTTGGCCTGGCCAAAGGCTTTCACAGC 340
QY 116 AGCAGAGAAAGTGAAGCTCAGACATACTGGCAGTGCACCAACAGACAGAGAACCTATG 175
Db 341 AGCAGGCGTCTTGGCTGCCACAGATATGGTAGTCCC--CAGACATGGACCTGGC 397
QY 176 TCACATTCAGCTC-----TGGGTTGCCGAGCAGCAGCTCCCAACAGGAGGCT 226
Db 398 CACAGTTTCTGCTCAGGAAGATGGGGTGAAGTTGACACCCCAACAAAGGAGCGGT 457
QY 227 GCAGAGCTGCTGCTGCTCATCTTTGCCACCAAGCCTCATGTGCTGCCAGCTGTCTGGC 286
Db 458 GCAGCAGTGTATGTGCTTCTGCTGTGAAGCCACACATATCCCTTCATCTCTGGA 517
QY 287 AGAGTGGCTCCTGTGTGCTACCACTGAACACATCTTGGTGTCCGTGGCTGTCTGGGGTGT 346
Db 518 TGAATAGGCGCGGACATTTGAGGACAGACACATTTGCTGTCTGCGGCGCGGCGTCCAC 577
QY 347 TCTGAGCACCTGGAGGAGCTGTG-----CCCCCAACACACAGGCTGTCTGGCGGT 397
Db 578 CATCAGCTCCATTGAGGAAGAGCTGTGAGCGCTTTCGGCCAGCCCCCAGGGTCTATCCGCTG 637
QY 398 CTTGCCCAACCTGCTGTGTGTCAGGAGGCGCCATAGTGTATGGCGGGGCGGCCCA 457
Db 638 CATGACCAACTCCAGTCTGGTGGGAGGCGGCCACCGTGTATGCCAGAGCGAC 697
QY 458 CGTGGGAGCAGCAGAGCAAGCTCTCTGCAGCATCTGTGAGGCGCTGTGGCGGTGTGA 517
Db 698 GCCCAGGTGGAG-GACGGGAGGCTCATGGAGCAGCTGTGAGCAGCGTGGGCTTCTGCAC 756
QY 518 GGAGTGGCTTGAAGCCTTACGTTCGACATCCACACTTGGGCTCAGTGGCAGTGGCGCTT 577
```

Db 757 GGAGGTGGAGAGGACCTGATTGATGCCCTACGGGGGCTCACTGGCAGCGGCCCGCCTA 816
QY 578 CBTGTGTGATTTCTCGAGGCGCTTGGCTGAAGGAGCCGTCAAGATGGGCGATGCCAGAG 637
Db 817 CGCATTCACAGCCCTGGATGGCTTGGCTGATGGGGTGTGAAGATGGGACTTCCAAGGCG 876
QY 638 CTTGGCCACCGGATCGCTGCCAGACCCCTGTGTGGGAGGCCAAGATGCTGCTGCACGA 697
Db 877 CTTGGCAGTCCGCTCGGGGCCAGGCGCTTCTGGGGGCTGCCAAGATGCTGCTGCACTC 936
QY 698 GGCCCAACACCCAGCCAGCTGGCTCAACACCTGTGCACCCCGGTGGCCACCATCTA 757
Db 937 AGACAGACCCAGCCAGCTCAAGACAACTGACGCTCTCTGTGGGGCCACCATCA 996
QY 758 TGGACTCCAGCCCTGGAGCAGGGGGGTGGAGCAGCCACCATGAGCGCGTGGAGGC 817
Db 997 TGCCTTGCATGTCTGGAGAGTGGGGCTTCCGCTCCCTGCTCATCAACGCTGTGGAGC 1056
QY 818 TGGCACTCGCGGGCCAGGAGCT 841
Db 1057 CTCTGCTATCCGCACACGGGAGCT 1080

RESULT 8

US-10-161-418A-4
; Sequence 4, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161,418A
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1028
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (999)..(999)
; OTHER INFORMATION: "n" is A, C, G, or T
US-10-161-418A-4

Query Match 15.8%; Score 211.2; DB 15; Length 1028;
Best Local Similarity 58.2%; Pred. No. 3.3e-42;
Matches 438; Conservative 0; Mismatches 293; Indels 21; Gaps 3;
QY 56 GCGCGTGGGCTTCGTGGCGCGGGCGCATGGCGGGGCGCATCGCAGGGGCTTCATCAG 115
Db 61 GAGCGTGGGCTTCATCGCGCTGGCCAGCTGGCTTTTGGCCCTGGCCAGGGCTTCACAGC 120
QY 116 AGCAGAAAGTGGAGCTCAGCATCTACTGGCCAGTGCACCAACAGACAGAACTATG 175
Db 121 AGCAGGCGTCTTGGCTGCCCAAGATAATGGTAGTCTCC---CAGACATGGAGCTGGC 177
QY 176 TCACCTTTCAAGTC-----TGGGTTGCCGACACGACACATCCCAACAGGAGTGCT 226
Db 178 CACAGTTTCTGCTCAGGAAGATGGGGTGAAGTTGACATCCCAACAGGAGAGCGGT 237
QY 227 GCAGAGTGCCTGCTGCTCATCTTTGCCAACCAAGCCCTCATGTGCTGCCAGCTGTCCTGGC 286
Db 238 GCAGCAGAGTAGTGTGCTCTCTCTGCTGTGAAGCCACACATATCCCTTCATCCCTGGA 297
QY 287 AGAGTGTGCTGCTGCTGCTCACCCTGACACATCTTGGTTCGCTGGCTGCTGGGCTGTC 346
Db 298 TGAATAGCGCCGACATTTAGGAGACAGACATTTGTGTGCTGCTGCGCGCGCGCGCTCAC 357

QY 347 TCTGAGCACCCTGGAGGAGCTGCTG-----CCCCCAACACACGGGTGCTGGGGT 397
Db 358 CATCAGCTCCATTGAGAAGAAGCTGTTCAGCGTTTCGCCACAGCCCCCAGGCTCATCCGCTG 417
QY 398 CTTGCCCAACCTGCCCTGTGTGTCCAGGAAGGGGCCATAGTGAATGGCGGGGGCGGCCA 457
Db 418 CATGACCAACACTCCAGTCTGTGTGGGAGGGGGCCACCGTGTATGCCACAGCAGCA 477
QY 458 CGTGGGAGCAGCAGACCAAGCTCCCTGCAGCATCTCTCTGGAGCCCTGTGGCGGCTGTGA 517
Db 478 CGCCAGGTGGAGGAGGGAGGCTCATGGAGCAGCTGCTGAGCAGCGTGGGCTTCTGCAC 537
QY 518 GGAGTGCCTGAAGCCCTACGTCGACATCCACACTGGCCCTCAGTGGCAGTGGCGCTT 577
Db 538 GGAGTGAAGAGGAGACCTGATTGATGCCGTACGGGGCTCAGTGGCAGCGCCCCGCCCTA 597
QY 578 CGTGTGTGCTATTTCCAGGCGCTTGAAGAGCGCGTCAAGATGGGCAATGCCAGCAG 637
Db 598 CGCATTCACAGCCCTGGATGCCCTGGTGGGGTGTGAAGATGGGACTTCCAAGCG 657
QY 638 CTTGGCCACCGCATCGCTGCCAGACCCCTCTGGGAGCGGCCAAGATGCTGTCACGA 697
Db 658 CTTGGCAGTCCGCTCGGGGCCAGGCCCTCTCTGGGGCTGCCAAGATGCTGTCACATC 717
QY 698 GGGCCAAACACCCAGCCAGCTGCGCTCAGACGTGTGCACCCGGGTGGCACCACCATCTA 757
Db 718 AGAAGCAGCACCAGCGCAGCTCAAGGACAACGTCAGCTCTCTGTGGTGGGGCACCATCCA 777
QY 758 TGGACTCCAGCCCTGGAGCAGGGCGGGCTGC 789
Db 778 TGCTTGCATGTCTGGAGAAAGTGGGGGCTTC 809

RESULT 9

US-09-925-300-278
; Sequence 278, Application US/09925300
; Patent No. US2002015181A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 278
; LENGTH: 1478
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-278

Query Match 15.2%; Score 203.4; DB 11; Length 1478;
Best Local Similarity 57.9%; Pred. No. 2.6e-40;
Matches 466; Conservative 0; Mismatches 316; Indels 23; Gaps 5;
QY 56 GCGCGTGGGCTTCGTGGCGCGGGCGCATGGCGGGGCGCATCGCAGGGGCTTCATCAG 115
Db 78 GAGCGTGGGCTTCATCGGCGCTGGCCAGCTGGCTTTTGGCCCTGGCCAAAGGGCTTCACAG 137
QY 116 -AGCAGAAAGTGGAGCTCAGCATCTACTGGCCAGTGCACCAACAGACAGAACTAT 174
Db 138 CAGCAGCGCTCTTGGCTGCCCAAGATAATGGTAGTCTCC---CAGACATGGAGCTGG 194
QY 175 GTCACCTTCAAGTC-----TGGGTTGCCGACACGACATCCCAACAGGAGGTGC 225
Db 195 CCACAGTTTCTGCTCTCAGGAAGATGGGGTGAAGTTGACATCCCAACAGGAGAGCG 254
QY 226 TGCAGAGCTGCCTGCTGCTCATCTTTGCCAACCAAGCCCTCATGTGCTGCCAGCTGCTCTGG 285

[illegible]

RESULT 10

US-10-161-418A-8 ; Sequence 8, Application US/10161418A
; Publication No. US20030036078A1

```

;
; GENERAL INFORMATION:
;
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRs AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161.418A
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1708
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-161-418A-8

```

Qy	102	CAGGGCCTCATAGAGCAGGAAGTGGAAAGCTCAGACATATCTGGCCAGTGCACCAACA	161
Db	204	CGGGGTTACGGCGCGCAGCATCTGTCTGGCTCACAAGATAATAGCAGCTCCCGAGNA	263
Qy	162	GACAGGNACC-----TATGTCACTTCAAGCTCTGGTTGCCGACCAGCACTCCAAC	215
Db	264	ATGAACCTGCCACGGGTCTCGGCGCTCAGGAAGATGGGTGTGAACCTGACACGAGCAAC	323
Qy	216	CAGGAGGTGTCGAGAGCTGCCTGCTCATCTTTTGCCACCAGCCTCATGTGCTGCCA	275
Db	324	AAGGAGAGCGTGAACACACGAGCGTCTCTTTCTGCTGTGAAGCCACATATCATCCCC	383
Qy	276	GCTGTCTCGCAGAGGTGGTCTCTGTGGTCAACACTGAACACATCTTTGGTGTCCGTGGCT	335
Db	384	TTCATCTCGATGAGATTTGGGGCGAGCTGCAAGCCAGACACATCGTGTGGTCTCTGTGGC	443
Qy	336	GCTGGGGTGTCTGAGCACCCCTGGAGAGCTGCTG-----CCCCAAACACACGG	386
Db	444	GCTGGTGTCAACATCAGCTCTGTGGAGAAGCTGATGGCANTTCCAGCCAGCCCCAAA	503
Qy	387	GTGCTGGGGTCTTGCCAAACCTGCCCTGTGTGTCAGGAAGGGGCCATAGTGAATGGCG	446
Db	504	GTGATTGCTGTCATGACCAACACACCTGTGTTAGTGCAGGAAGCGCTACAGTGTACGCC	563
Qy	447	CGGGGCGCCAGTGGGGAGGAGCGGAGACCAAGTCTCTGCAGCATCTGCTGGAGGCGCTGT	506
Db	564	ACGGGCACCATGCGCTTGGTGGAGATGGGCAGCTCCTGGAGCAGCTCATGAGCAGCGTG	623
Qy	507	GGGCGGTGTGAGGAGTGCCTGGAAGCCTACGTCCACATCCACACTGCCCTCAGTGGCAGT	566
Db	624	GGCTTCTGCATGAGGTGGGAAGAGGACCTCATCGATGCCGTTCAGGGGCTCAGTGGCAGC	683
Qy	567	GGCTGGCCTTCTGTGTGTGCATTCTCGAGGCCCTGGCTGAAGGAGCGCTCAAGATGGGC	626
Db	684	GGGCTTGCCTATGATTTCATGGCTCTGGACGCAATTTGGCTGATGTTGGGTGAAGATGGT	743
Qy	627	ATGCCACGACGCTTGGCCACCACCGCATCGTCCCGACAGCCCTGCTGGGGAGCGGCCAAGATG	686
Db	744	TTGCCAGGGCGCCTTGGCAATCCAACCTCGGGGCCCAGGGCTTTGCTGGGAGCTTGGCAAGATG	803
Qy	687	CTGCTGCAGAGGGCCACACCCAGCCAGCTGCGCTCAGACGTTGTCACCCCGGGTGGC	746
Db	804	CTGTGGACTCGGAGCAGCATCCATGCCAGCTTAAGGACAATGTCTGTCTCCCTCGGGGGA	863
Qy	747	ACCACCATCTATGGACTCCAGCCCTTGGAGCAGGGCGGGCTGCGAGCAGCCACCATGAGC	806
Db	864	GCCACCATCCAGCCCTGCACCTTTCATAGAGAGTGGGGCTTCCCGCTCTCTGCTCATCAAT	923
Qy	807	GCCGTGGAGGTGCCACCTCGCGGGCCCAAGGAGCT	841
Db	924	GCAGTTGAGGCTCTGTATCCGAACACGAGAGCT	958

RESULT 11

US-09-912-717-2
; Sequence 2, Application US/09912717
; Patent No. US20020081691A1

GENERAL INFORMATION:
 APPLICANT: Hillman, Jennifer L.
 Corley, Neil C.
 Baughn, Mariah R.
 TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
 HOMOLOG
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 04:52:43 ; Search time 2117 Seconds
(without alignments)
10235.976 Million cell updates/sec

Title: US-09-806-536A-29
Perfect score: 1338
Sequence: 1 ggtgagcgcagtctgtccga.....ggggcacgagcatatgggggt 1338

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_fod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	884	66.1	1086	13	BM548059
2	871	65.1	906	9	AL526237
3	859.2	64.2	885	9	AL519338
4	832.2	62.2	1080	14	BM923064
5	786.4	58.8	873	13	BI837148
6	785.6	58.7	797	9	AL525111

7	763.8	57.1	773	13	BI827333	BI827333	603077811
8	753	56.3	856	13	BM014974	BM014974	603640923
9	737.6	55.1	794	9	AU141625	AU141625	603640923
10	701.2	52.4	1145	13	BM473687	BM473687	AGENCOURT
11	697.4	52.1	932	14	BQ436653	BQ436653	AGENCOURT
12	688.2	51.4	734	13	BI253559	BI253559	602973485
13	673.6	50.3	697	9	AL522085	AL522085	602973485
14	660	49.3	741	13	BM010281	BM010281	603631154
15	627.6	46.9	773	13	BI818555	BI818555	603033078
16	622.4	46.5	664	9	AI963473	AI963473	603033078
17	620.4	46.4	679	12	AG393676	AG393676	602412094
18	617	46.1	895	12	BF797243	BF797243	602257745
19	614.4	45.9	1036	13	BM545513	BM545513	AGENCOURT
20	589.8	44.1	1024	13	BM465183	BM465183	AGENCOURT
21	580.4	43.4	590	12	AG282890	AG282890	602406016
22	569.2	42.5	1271	11	AK002912	AK002912	Mus muscu
23	567.8	42.4	823	13	BM019985	BM019985	603648507
24	557.4	41.7	595	10	BE621964	BE621964	601440609
25	552.8	41.3	1132	13	BM424026	BM424026	AGENCOURT
26	535.8	40.0	1265	11	AK013214	AK013214	Mus muscu
27	535	40.0	803	14	BQ571032	BQ571032	UI-M-FBO-
28	532	39.8	744	10	BE267359	BE267359	601189558
29	532	39.8	902	14	BQ964436	BQ964436	AGENCOURT
30	527.6	39.4	788	13	BI254417	BI254417	602976749
31	525.6	39.3	874	13	BI415573	BI415573	602988350
32	520.8	38.9	1348	11	AK004291	AK004291	Mus muscu
33	520.6	38.9	737	13	BM019007	BM019007	603646981
34	517.4	38.7	670	10	BE563597	BE563597	601334712
35	510.6	38.2	869	13	BI329060	BI329060	602980475
36	505.2	37.8	602	9	AL706316	AL706316	DRF2P686C
37	504.4	37.7	1178	13	BM461077	BM461077	AGENCOURT
38	495.6	37.0	722	14	BM945806	BM945806	UI-M-EMO-
39	486.6	36.4	563	9	AU129509	AU129509	60129509
40	482.6	36.1	752	14	BM945714	BM945714	UI-M-EMO-
41	465.4	34.8	736	13	BI903838	BI903838	603166049
42	457.8	34.2	476	9	AL556719	AL556719	603166049
43	453.6	33.9	660	14	BQ571218	BQ571218	UI-M-FBO-
44	449	33.6	741	13	BG967901	BG967901	602833147
45	430.8	32.2	619	12	BF787218	BF787218	602108978

ALIGNMENTS

RESULT 1
BM548059 1086 bp mRNA linear EST 20-FEB-2002
LOCUS AGENCOURT_6531538 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5732882
DEFINITION 5', mRNA sequence.
ACCESSION BM548059.1 GI:18782344
VERSION BM548059
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1086)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM12735 row: 1 column: 03
High quality sequence start: 13
High quality sequence stop: 685.
Location/Qualifiers

FEATURES


```
|||||
120 ACTGGCAGTGCACCAACAGACAGAACTATGTCACTTCAAGCTCTGGGTGCGGAC 179
QY 203 CAGGCACCTCAACACGAGAGTCTGCAGAGTCTGCTGCTGCTCACTTTTGCACCAAGCC 262
Db 180 CAGGCACCTCAACACGAGAGTCTGCAGAGTCTGCTGCTCACTTTTGCACCAAGCC 239
QY 263 TCATGTCTGCAGCTGCTCTGGCAGAGTGTGCTCTGCTGCTCACTTGAACACATCTT 322
Db 240 TCATGTCTGCCAGCTGCTCTGGCAGAGTGTGCTCTGCTGCTCACTTGAACACATCTT 299
QY 323 GTGTCCCGTGTCTGGGTCTCTTGAGCACCCTTGAGGAGTGTCTGCCCCCAACAC 382
Db 300 GTGTCCCGTGTCTGGGTCTCTTGAGCACCCTTGAGGAGTGTCTGCCCCCAACAC 359
QY 383 ACGGTGTCTGCGGGTCTTGCACCACTGCCCTGTGTGTCTCCAGGAAGGGCCATAGTAT 442
Db 360 ACGGTGTCTGCGGGTCTTGCACCACTGCCCTGTGTGTCTCCAGGAAGGGCCATAGTAT 419
QY 443 GCGGGGGCCCGCAGTGGGAGCAGCAGACCAAGCTCTCTGACGATCTCTGCGAGGC 502
Db 420 GCGGGGGCCCGCAGTGGGAGCAGCAGACCAAGCTCTCTGACGATCTCTGCGAGGC 479
QY 503 CTGTGGCGGTGTGAGGAGTGCCTGAAGCTTACGTTCGACATCCACACTGGCCTCAGTGG 562
Db 480 CTGTGGCGGTGTGAGGAGTGCCTGAAGCTTACGTTCGACATCCACACTGGCCTCAGTGG 539
QY 563 CAGTGGCGTGGCTTGTGTGTGATTTCCGAGGCCCTGGCTGAAGAGCCGCTCAAGAT 622
Db 540 CAGTGGCGTGGCTTGTGTGTGATTTCCGAGGCCCTGGCTGAAGAGCCGCTCAAGAT 599
QY 623 GGGCATGCCACAGAGCTGGCCACCGCATCGCTGCCAGACCCCTGCTGGGACGCCAA 682
Db 600 GGGCATGCCACAGAGCTGGCCACCGCATCGCTGCCAGACCCCTGCTGGGACGCCAA 659
QY 683 GATGCTGCTGCAGAGGGCCAAACACCCAGCCAGCTGCGCTCAGAGCTGTGACCCCGGG 742
Db 660 GATGCTGCTGCAGAGGGCCAAACACCCAGCCAGCTGCGCTCAGAGCTGTGACCCCGGG 719
QY 743 TGCACACATATATGAGTCCAGCCCTTGAGCAGGCGGGCTGCGAGCAGCCACCAT 802
Db 720 TGCACACATATATGAGTCCAGCCCTTGAGCAGGCGGGCTGCGAGCAG -CACCAT 778
QY 803 GAGCGCGTGGAGCTGCCACCTGCGGGCCAAAGAGCTCAGCAGAAAGTAGGCTGGCT 862
Db 779 GAGCGCGTGGAGCTGCCACCTGCGGGCCAAAGAGCTCAGCAGAAAGTAGGCTGGCT 838
QY 863 CTGGCCATCCTTTCTGCTGCTGTGTGCCCTGCTCTCCCTGTGTGCTTCCCTGAGGAC 922
Db 839 CTGGCCATCCTTTCTGCTGCTGTGTGCCCTGCTCTCCCTGTGTGCTT -CCCTGAAGAC 897
QY 923 TGGGGCTCC 931
Db 898 TGGGGCTCC 906

RESULT 3
LOCUS AL519338 885 bp mRNA linear EST 13-FEB-2001
DEFINITION AL519338 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB004Yf19 5
prime, mRNA sequence.
ACCESSION AL519338
VERSION AL519338.1 GI:12782831
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 885)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
```

```
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB004Yf19"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 147 a 290 c 156 t 2 others
ORIGIN
Query Match 64.2%; Score 859.2; DB 9; Length 885;
Best Local Similarity 99.5%; Pred. No. 2.1e-183;
Matches 880; Conservative 2; Mismatches 0; Indels 2; Gaps 2;
QY 31 TGGCAGCTGCGGAGCCGCTCCCGGGCGCTGGGCTGTGGTGGCGGGCGCGCATGGCGG 90
Db 1 TGGCAGCTGCGGA -CCGTCTCCGGCGCGCTGGGCTGTGGTGGCGGGCGCGCATGGCGG 59
QY 91 GGGCCATCGCCAGGGCCCTCATCAGCAGCAAGTGGAAAGTGAAGCTCAGCACATACTGGCCA 150
Db 60 GGGCCATCGCCAGGGCCCTCATCAGCAGCAAGTGGAAAGTGAAGCTCAGCACATACTGGCCA 119
QY 151 GTGCACCAACAGACAGGAACCTATGTCACTTTCAAGCTCTGGTGGCGGACCACGCACT 210
Db 120 GTGCACCAACAGACAGGAACCTATGTCACTTTCAAGCTCTGGTGGCGGACCACGCACT 179
QY 211 CCAACGAGGAGTGTGTCAGAGTGCCTGCTGCTGCTATCTTTGCCACCAAGCCTCATGTGC 270
Db 180 CCAACGAGGAGTGTGTCAGAGTGCCTGCTGCTGCTATCTTTGCCACCAAGCCTCATGTGC 239
QY 271 TGCCAGCTGCTCTGTCAGAGTGGCTCTGCTGCTGCTCACCCTGACACATCTTGGTGTCCG 330
Db 240 TGCCAGCTGCTCTGTCAGAGTGGCTCTGCTGCTGCTCACCCTGACACATCTTGGTGTCCG 299
QY 331 TGGCTGCTGGGTGTCTCTGAGCACCCTTGAGGAGTCTGCTGCCCCCAACACAGGGTGC 390
Db 300 TGGCTGCTGGGTGTCTCTGAGCACCCTTGAGGAGTCTGCTGCCCCCAACACAGGGTGC 359
QY 391 TGGGGTCTTGCCCAACCTGCCCTGTGTGCTCAGGAAGGGCCATAGTATGGCGGG 450
Db 360 TGGGGTCTTGCCCAACCTGCCCTGTGTGCTCAGGAAGGGCCATAGTATGGCGGG 419
QY 451 GCGGCAGCTGGGGAGCAGCAGACCAAGCTCTCTGACATCTGCTGGAGGCTGTGGGC 510
Db 420 GCGGCAGCTGGGGAGCAGCAGACCAAGCTCTCTGACATCTGCTGGAGGCTGTGGGC 479
QY 511 GGTGTGAGGAGTGTCTGAAGCCTACGTGACATCCACACTGGCTCAGTGGCAGTGGCG 570
Db 480 GGTGTGAGGAGTGTCTGAAGCCTACGTGACATCCACACTGGCTCAGTGGCAGTGGCG 539
QY 571 TGGCCTTGTGTGCTGCTTCCGAGGCCCTGGCTGAAGGAGCCCTCAAGATGGCATGC 630
Db 540 TGGCCTTGTGTGCTGCTTCCGAGGCCCTGGCTGAAGGAGCCCTCAAGATGGCATGC 599
QY 631 CCAGCAGCTGGCCCCACCGCATCTGCTGCCAGACCCCTGCTGGGGACGCCCAAGATGCTGC 690
Db 600 CCAGCAGCTGGCCCCACCGCATCTGCTGCCAGACCCCTGCTGGGGACGCCCAAGATGCTGC 659
```

```
FEATURES
source
1. 885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB004Yf19"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
```


QY 691 TGCACGAGGCGCCAAACACCCAGCCAGCTGCGCTCAGACGTTGTGACACCCCGGGTGGCACCA 750
|||||
Db 660 TGCACGAGGCGCCAAACACCCAGCCAGCTGCGCTCAGACGTTGTGACACCCCGGGTGGCACCA 719
|||||
QY 751 CCATCTATGACATCCAGCGCTCGAGCAGGGCGGGTGGAGCAGCCACCATGAGCGCG 810
|||||
Db 720 CCATCTATGACATCCAGCGCTCGAGCAGGGCGGGTGGAGCAGCCACCATGAGCGCG 779
|||||
QY 811 TGGAGGCTGCACCTCGCGGGCCCAAGAGCTCAGCAGAAAGTAGGCTGGGCTCTGGCCAT 870
|||||
Db 780 TGGAGGCTGCACCTCGCGGG-OMAGAGCTCAGCAGAAAGTAGGCTGGGCTCTGGCCAT 838
|||||
QY 871 CCTTCTGCTCTGTGCGCCCTGCCCTCTCCCTGTGTCCCTCCC 914
|||||
Db 839 CCTTCTGCTCTGTGCGCCCTGCCCTCTCCCTGTGTCCCTCCC 882
|||||

RESULT 4

BM923064

LOCUS

DEFINITION BM923064 1080 bp mRNA linear EST 12-MAR-2002
AGENCOURT_6631985 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5757031
5', mRNA sequence.

ACCESSION

BM923064

VERSION

BM923064.1

GI:19373443

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1080)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM12798 row: j column: 08

High quality sequence stop: 609.

Location/Qualifiers

1..1080

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5757031"

/clone_lib="NIH_MGC_118"

/tissue_type="Leukocyte"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV

(destroyed); RNA source leukocytes from anonymous pool of

non-activated adult donors. Library is oligo-dT primed

and directionally cloned (EcoRV site is destroyed upon

cloning). Average insert size 1.7 kb, insert size range

1.2-3.3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(invitrogen). Research Genetics tracking code 027. Note:

this is a NIH_MGC Library."

BASE COUNT 189 a 358 c 350 g 183 t

ORIGIN

Query Match

Best Local Similarity 62.2%; Score 832.2; DB 14; Length 1080;

Matches 973; Conservative 91.7%; Pred. No. 2.9e-177;

Mismatches 68; Indels 20; Gaps 8;

QY 15 GTCCAGGCAACAGATGCGAGCGTCTCGCGCGCGGTGGGCTCTGGGC 74

|||||

Db 15 GTCCAGGCAAC-AGATGGAGCTGGGAGCGCTCTCGCGCGCGGTGGGCTCTGGGC 73

|||||

QY 75 GCGGGCGCGATGGCGGGGGCCATCGCGCAGGGCCCTCATCAGAGCAGGAAAGTGGAGCT 134

Db 74 GCGGGCGCGATGGCGGGGCCCATCGCGAGGGCCCTCATCAGAGCAAGAAAGTGGAGCT 133
|||||
QY 135 CAGCATACTATGGCCAGTGCACCAACAGACAGAGAACCTATGTCACTTTCAAGCTCTGGGT 194
|||||
Db 134 CAGCATACTATGGCCAGTGCACCAACAGACAGAGAACCTATGTCACTTTCAAGCTCTGGGT 193
|||||
QY 195 TGGCGGACACGCACTCCAAACAGGAGGTGCTGCAGAGCTGCCCTGCTCATCTTTTGGC 254
|||||
Db 194 TGGCGGACACGCACTCCAAACAGGAGGTGCTGCAGAGCTGCCCTGCTCATCTTTTGGC 253
|||||
QY 255 ACCAAGCCTCATGTGCTGCCAGCTGTCTTGCACAGAGTGGCTCTGTGGTCAACACCTGAA 314
|||||
Db 254 ACCAAGCCTCATGTGCTGCCAGCTGTCTTGCACAGAGTGGCTCTGTGGTCAACACCTGAA 313
|||||
QY 315 CACATCTTGGTGTCCGTGGGTGTCTGTAGACACCTTGGAGGAGCTGTCTGCC 374
|||||
Db 314 CACATCTTGGTGTCCGTGGGTGTCTGTAGACACCTTGGAGGAGCTGTCTGCC 373
|||||
QY 375 CCAACACACGGGTGCTGCGGGTCTTGCACCACTGCCCTGTGTGGTCCAGGAGGGGCC 434
|||||
Db 374 CCAACACACGGGTGCTGCGGGTCTTGCACCACTGCCCTGTGTGGTCCAGGAGGGGCC 433
|||||
QY 435 ATAGTGATGGCGCGGGGCCCGCCACGTGGGAGCAGCAGCAAGCTCTCTGCAGCATCTG 494
|||||
Db 434 ATAGTGATGGCGCGGGGCCCGCCACGTGGGAGCAGCAGCAAGCTCTCTGCAGCATCTG 493
|||||
QY 495 CTGAGAGCCTGTGGCGGTGTGAGGAGGTGCTTGAAGCTTACGTGCACATCCACACTGGC 554
|||||
Db 494 CTGAGAGCCTGTGGCGGTGTGAGGAGGTGCTTGAAGCTTACGTGCACATCCACACTGGC 553
|||||
QY 555 CTCAGTGGCAGTGGCGTGGCTTTCGTGTGTGCATTTCCGAGGCGCTTGGCTGAAGAGCC 614
|||||
Db 554 CTCAGTGGCAGTGGCGTGGCTTTCGTGTGTGCATTTCCGAGGCGCTTGGCTGAAGAGCC 613
|||||
QY 615 GTCAGATGGCATGGCCAGCAGCTGGCCACCGCATCGCTGCCAGACCTCTGCTGGGG 674
|||||
Db 614 GTCAGATGGCATGGCCAGCAGCTGGCCACCGCATCGCTGCCAGACCTCTGCTGGGG 673
|||||
QY 675 ACGCCAAGATGCTGTGTCAGAGGGCCAAACCCAGCCAGCTGGCTCAGACGCTGTC 734
|||||
Db 674 ACGCCAAGATGCTGTGTCAGAGGGCCAAACCCAGCCAGCTGGCTCAGACGCTGTC 733
|||||
QY 735 ACCCGGGTGGCA-CCACCATCTATGACTTCCACGCGCTTGGAGCAGGGCGGCTGCGAGC 793
|||||
Db 734 ACCCGGGTGGCA-CCACCATCTATGACTTCCACGCGCTTGGAGCAGGGCGGCTGCGAGC 793
|||||
QY 794 AGCCACCATGAGCG-CGGTGGAGCTGCCACCTTCCGCGGCCCAAGAGCTCAGCAGAAAGT 852
|||||
Db 794 AGCCACCATGAGCGCTTGGAGCTGCCACCTTCCGCGGCCCAAGAGCTCAGCAGAAAG 853
|||||
QY 853 AGGCTGGGCTCTGGCCATCCCTTTCCCTCTCTGTGCCCTGCTCTCTCTCTCTCTCTCTCT 909
|||||
Db 854 TAGCTGGGCTCTGGGCATCCCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 913
|||||
QY 910 TTCCCTTGAGGACT--GCGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 965
|||||
Db 914 TTCCCTTGAGGACTTGGGGGTCT 973
|||||
QY 966 CTCCTCTTGACAGGAGGAAA---TGCAGGGGGCAGCAGCTTTGGAGGCTT-----CCAGCA 1015
|||||
Db 974 CTCCTCTTGACAGGAGGAAA---TGCAGGGGGCAGCAGCTTTGGAGGCTT-----CCAGCA 1015
|||||
QY 1016 GCGGGGAGGAGCCCGCCAGCTGGGAGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1056
|||||
Db 1034 GCGGGGAGAGCCCGCCAGCTGGGAGAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1074
|||||

RESULT 5

BI837148

LOCUS

DEFINITION

603089920F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5228956 5',

mRNA sequence.

BI837148 873 bp mRNA linear EST 04-OCT-2001
603089920F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5228956 5',
mRNA sequence.

ORIGIN

Query Match	58.7%	Score 785.6	DB 9	Length 797
Best Local Similarity	99.7%	Pred. No. 8e-167		
Matches 796	Conservative 1	Mismatches 1	Indels 0	Gaps 1
QY	25	ACAAGATGGCAGCTGCGAGCGCTCCGCGGCGCTTCGTGGCGCGCGGCCCA	84	
Db	1	ACAAGATGGCAGCTGCGGA-SCGTCTCCGCGCGCTGGCTTCGTGGCGCGGCCCA	59	
QY	85	TGGCGGGGCCATCGCGAGGCGCTCATCAGACAGAGAAAGTGAAGCTCAGCACATAC	144	
Db	60	TGGCGGGGCCATCGCGAGGCGCTCATCAGACAGAGAAAGTGAAGCTCAGCACATAC	119	
QY	145	TGSCCAGTGCACCAACACAGACAGAAACCTATGTCACTTTCAAGCTCTGGGTTGCCGGACCA	204	
Db	120	TGSCCAGTGCACCAACACAGAAACCTATGTCACTTTCAAGCTCTGGGTTGCCGGACCA	177	
QY	205	CGCACTCCAACAGGAGGTGCTGCAGAGCTGCCTGCTCATCTTTGCCACCAAGCGCTC	264	
Db	180	CGCACTCCAACAGGAGGTGCTGCAGAGCTGCCTGCTCATCTTTGCCACCAAGCGCTC	239	
QY	265	ATGTGCTGCCAGCTGTCTTGGCAGAGGTGGCTCCTGTGTCAACACTGAACACATCTTGG	324	
Db	240	ATGTGCTGCCAGCTGTCTTGGCAGAGGTGGCTCCTGTGTCAACACTGAACACATCTTGG	299	
QY	325	TGTCGGTGGCTGCTGGGGTGTCTTGAGACACCTGGAGGAGCTGCTGCCCCCAACACAC	384	
Db	300	TGTCGGTGGCTGCTGGGGTGTCTTGAGACACCTGGAGGAGCTGCTGCCCCCAACACAC	359	
QY	385	GGTGCTCGGGGTCTTGCCCAACTGCCCTGTGTGTCCAGGAAGGGCCATAGTGATGG	444	
Db	360	GGTGCTCGGGGTCTTGCCCAACTGCCCTGTGTGTCCAGGAAGGGCCATAGTGATGG	419	
QY	445	CGCGGGGCGCCAGCTGGGAGCAGGACCAAGCTCTGCAGCATCTGCTGGAGCGCT	504	
Db	420	CGCGGGGCGCCAGCTGGGAGCAGGACCAAGCTCTGCAGCATCTGCTGGAGCGCT	479	
QY	505	GTGGCGGTGTGAGGAGGTGCCTGAAGCCTAGCTGCACATCCACACTGGCTCAGTGCCA	564	
Db	480	GTGGCGGTGTGAGGAGGTGCCTGAAGCCTAGCTGCACATCCACACTGGCTCAGTGCCA	539	
QY	565	GTGGCTGGCTTCGTGTGCAATCTCCGAGGCGCTGGCTCAAGAGCCGCTCAAGATGG	624	
Db	540	GTGGCTGGCTTCGTGTGCAATCTCCGAGGCGCTGGCTCAAGAGCCGCTCAAGATGG	599	
QY	625	GCATGCCACAGCCTGGGCCACCGCATCGCTGCCACAGCCCTGCTGGGAGCGCCAAGA	684	
Db	600	GCATGCCACAGCCTGGGCCACCGCATCGCTGCCACAGCCCTGCTGGGAGCGCCAAGA	659	
QY	685	TGCTGCTGCAGAGGGCCACACCCAGCCAGCTGGGTTCAGACGTGTGCACCCCGGTG	744	
Db	660	TGCTGCTGCAGAGGGCCACACCCAGCCAGCTGGGTTCAGACGTGTGCACCCCGGTG	719	
QY	745	GCACCACTATATGACTCCAGCCCTCGAGCGGGGCTGCGACGACGCCACCATGA	804	
Db	720	GCACCACTATATGACTCCAGCCCTCGAGCGGGGCTGCGACGACGCCACCATGA	779	
QY	805	CGCGCGTGGAGGCTGCCA	822	
Db	780	CGCGCGTGGAGGCTGCCA	797	

RESULT 7	
BI827333	
LOCUS	
DEFINITION	773 bp mRNA linear EST 04-OCT-2001 603077811F1 NIH_MGC_119 Homo sapiens CDNA clone IMAGE:5169410 5', mRNA sequence.
ACCESSION	BI827333
VERSION	BI827333.1 GI:15938883
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 773)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

FEATURES
SOURCE

```

ftp://image.dnli.90v
Plate: LLAM1421 row: f column: 03
High quality sequence stop: 766.
Location/Qualifiers
1. .773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5169410"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/notes="Organ: Brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dr primed and
directionally cloned [EcoRV site is destroyed upon
cloning]. Average insert size 1.3 kb, insert size range
0.9-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 013. Note:
this is a NIH_MGC Library."
138 a 246 c 258 g 131 t

```

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

Query Match	57.1%;	Score 763.8;	DB 13;	Length 773;
Best Local Similarity	99.7%;	Pred. No. 6.6e-162;		
Matches 765;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0
QY	18	CGAGGCAACAAGATGGCAGCTGGCGAGCGCTCCGCGCGCGCTGGCGTTCGTGGCGCGG	77	
Db	7	CGAGGCAACAAGATGGCAGCTGGCGAGCGCTCCGCGCGCGCTGGCGTTCGTGGCGCGG	66	
QY	78	GGCCGCAATGGCGGGGGCCCATCGCGAGGGCTCATCAGACGAGAAAGTGGGAAGCTCAG	137	
Db	67	GGCCGCAATGGCGGGGGCCCATCGCGAGGGCTCATCAGACGAGAAAGTGGGAAGCTCAG	126	
QY	138	CACATACTGGCCAGTGGCACCAACACAGAGGAACCTATGTCACCTTTTCAAGCTCTGGGTTCG	197	
Db	127	CACATACTGGCCAGTGGCACCAACACAGAGGACCTATGTCACCTTTTCAAGCTCTGGGTTCG	186	
QY	198	CGGACCAACGCACTCCAAACGAGAGGTGCTCGAGAGTGCCTGCTGCTCATCTTTTGGCAAC	257	
Db	187	CGGACCAACGCACTCCAAACGAGAGGTGCTCGAGAGTGCCTGCTGCTCATCTTTTGGCAAC	246	
QY	258	AAGCCTCATGTGCTGCCAGGTGCTCGGACAGAGTGGCTCCTCTGGTGCACCACTGAACAC	317	
Db	247	AAGCCTCATGTGCTGCCAGGTGCTCGGACAGAGTGGCTCCTCTGGTGCACCACTGAACAC	306	
QY	318	ATCTTTGGTGTCCGTGGCTGTGGGGTGTCTGTGACCAACCTGGAGAGCTGTGCCCCCA	377	
Db	307	ATCTTTGGTGTCCGTGGCTGTGGGGTGTCTGTGACCAACCTGGAGAGCTGTGCCCCCA	366	
QY	378	AACACACGGGTGCTGGCGGTCTTGCCCAACCTGCCCTGTGTGCTGCAGGAAGGGGCCATA	437	
Db	367	AACACACGGGTGCTGGCGGTCTTGCCCAACCTGCCCTGTGTGCTGCAGGAAGGGGCCATA	426	
QY	438	GTGATGGCGGGGGCCGCCACGTGGGAGACGAGACCAAGCTCTCGAGCATCTGTGTG	497	
Db	427	GTGATGGCGGGGGCCGCCACGTGGGAGACGAGACCAAGCTCTCGAGCATCTGTGTG	486	

```

QY 498 GAGCGCTGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 557
    |||
Db 487 GAGCGCTGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 546
    |||
QY 558 AGTGGCAGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 617
    |||
Db 547 AGTGGCAGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 606
    |||
QY 618 AAGATGGCAGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 677
    |||
Db 607 AAGATGGCAGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 666
    |||
QY 678 GCCAAGATGGCTGTCAGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 737
    |||
Db 667 GCCAAGATGGCTGTCAGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 726
    |||
QY 738 CCGGCTGGCAGCACCACCATCTATGAGTCCACGCCCTGGAGCAGGCGG 784
    |||
Db 727 CCGGCTGGCAGCACCACCATCTATGAGTCCACGCCCTGGAGCAGGCGG 773
    |||

RESULT 8
BM014974
LOCUS 603640923F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5417260 5',
DEFINITION mRNA sequence.
ACCESSION BM014974.1 GI:16529328
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs@mail.nih.gov
Tissue Procurement: DCFD/DFP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12064 row: m column: 05
High quality sequence stop: 783.
Location/Qualifiers
1. 856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5417260"
/clone_lib="NIH_MGC_87"
/lab_host="mammary adenocarcinoma, cell line"
/note="organ: breast; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 158 a 267 c 290 g 141 t
ORIGIN

Query Match 56.3%; Score 753; DB 13; Length 856;
Best Local Similarity 97.9%; Pred. No. 1.9e-159;
Matches 837; Conservative 0; Mismatches 10; Indels 8; Gaps 7;

QY 18 CGAGGCAACAAGATGCGAGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 77
    |||
Db 2 CGAGGCAACAAGATGCGAGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 61
    |||
QY 78 GCGCGCATGGCGGGGCCATCGCGCAGGCGCTCATCAGAGCAGGAGAAAGTGAAGCTCAG 137
    |||

```

```

    |||
Db 62 GCGCGCATGGCGGGGCCATCGCGCA - GCGCTCATCAGAGGAGAAAGTGAAGCTCAG 120
    |||
QY 138 CACATCTGCGCAGTGCACCAACAGACAGAGAACCTATGTCTACTTTCAAGCTCTGGGTGC 197
    |||
Db 121 CACATCTGCGCAGTGCACCAACAGACAGAGAACCTATGTCTACTTTCAAGCTCTGGGTGC 180
    |||
QY 198 CGSACCACGACCTCCAAACAGGAGGTGCTGCAGAGTGCCTGCTGCTCATCTTTGCGCAC 257
    |||
Db 181 CGSACCACGACCTCCAAACAGGAGGTGCTGCAGAGTGCCTGCTGCTCATCTTTGCGCAC 240
    |||
QY 258 AAGCCTCATGTGCTGCGCAGTGTCTTGGCAGAGTGGCTCTGTGTGCTACCACTGAACAC 317
    |||
Db 241 AAGCCTCATGTGCTGCGCAGTGTCTTGGCAGAGTGGCTCTGTGTGCTACCACTGAACAC 300
    |||
QY 318 ATCTTGGTGTGCTGCGTGGGTGTCTCTGAGCAGCCTTGAGAGAGTGTGCTGCCCCCA 377
    |||
Db 301 ATCTTGGTGTGCTGCGTGGGTGTCTCTGAGCAGCCTTGAGAGAGTGTGCTGCCCCCA 360
    |||
QY 378 AACACACGGGTGCTGCGGTGTCTTGGCCAAACCTGCCCTGTGTGCTCCAGGAAGGGGCCATA 437
    |||
Db 361 AACACACGGGTGCTGCGGTGTCTTGGCCAAACCTGCCCTGTGTGCTCCAGGAAGGGGCCATA 420
    |||
QY 438 GTGATGCGCGCGGGGCCACACGTGGGAGAGCAGACCAAGCTCTCTGAGCATCTGTGTG 497
    |||
Db 421 GTGATGCGCGCGGGGCCACACGTGGGAGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 480
    |||
QY 498 GAGGCGCTGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 557
    |||
Db 481 GAGGCGCTGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCCTC 540
    |||
QY 558 AGTGGCAGTGGCGCTTGTGTGTGCTTGTGCTTCCGAGGCGCTGTGTGCTGAAGGAGCGCTC 617
    |||
Db 541 AGTGGCAGTGGCGCTTGTGTGTGCTTGTGCTTCCGAGGCGCTGTGTGCTGAAGGAGCGCTC 600
    |||
QY 618 AAGATGGGATGCCCCAGCAGCCTTGCCCCACCGCATCGCTGCCAGACCCCTGTGCTGGGAGC 677
    |||
Db 601 AAGATGGGATGCCCCAGCAGCCTTGCCCCACCGCATCGCTGCCAGACCCCTGTGCTGGGAGC 659
    |||
QY 678 GCCAAGATGCTGTGTCAGAGGCGCCACACCCAGCCTACGTGCCTTCCAGCTGTGCA - AGACCTGTGCA - 735
    |||
Db 660 GCAAAGATGCTGTGTCAGAGGCGCCACACCCAGGCGCAGCTGCGCTCAAGCATGTGTGCAC 719
    |||
QY 736 CCGCGGTGGCACCACCATCTAT - GGACTCCAGCCTCTGGAGCAGGCG - - GGCCTGCGAG 792
    |||
Db 720 CCGCGGTGGCACCACCATCTATGAGTCCAGCCTCTGGAGCAGGCGCGGCTTGGAG 779
    |||
QY 793 CAGCACCATGAGCGCGTGGAGCTGCCACCTGCCCGGCGCAAGGAGCTCAGCAGAAAGT 852
    |||
Db 780 CAGCAACCATGAGCGCGGAGGCTGCAACCTGCCCGG - CAAAGAGGTGAGCAGAAAGT 838
    |||
QY 853 AGGCTGGCTCTGCG 867
    |||
Db 839 AGGCTGGCTCTGCG 853
    |||

RESULT 9
BM0141625
LOCUS THYR01 Homo sapiens cDNA clone THYR01000934 5', mRNA
DEFINITION sequence.
ACCESSION BM0141625
VERSION AUI141625.1 GI:11003146
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 794)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
TITLE

```

JOURNAL
COMMENT

Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES
source

Location/Qualifiers
1. .794
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01000934"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
142 a 248 c 264 g 133 t 7 others

Query Match 55.1%; Score 737.6; DB 9; Length 794;
Best Local Similarity 97.7%; Pred. No. 5.5e-156;
Matches 765; Conservative 0; Mismatches 16; Indels 2; Gaps 2;

QY 15 GTCGAGGCAACAAGATGTCAGTCGGAGCGGTCTCCGCGCGGTGGCTTCGTGGGC 74
DB 12 GTCGAGGCAACAAGATGTCAGTCGGAGCGGTCTCCGCGCGGTGGCTTCGTGGGC 71
QY 75 GCGGGCGCATGCGGGGGCCATCGCGCAGGCGCTCATCAGACGAGAAAGTGAAGCT 134
DB 72 GCGGGCGCATGCGGGGGCCATCGCGCAGGCGCTCATCAGACGAGAAAGTGAAGCT 131
QY 135 CAGCACAATCTGGCCAGTGCACCAACAGACAGAGAACCTATGTCATTTCAAGCTCTGGGT 194
DB 132 CAGCACAATCTGGCCAGTGCACCAACAGACAGAGAACCTATGTCATTTCAAGCTCTGGGT 191
QY 195 TGGCGGACACGCACTCCAACGAGGAGTGTGTCAGAGTGCCTGCTCATCTTTGCC 254
DB 192 TGGCGGACACGCACTCCAACGAGGAGTGTGTCAGAGTGCCTGCTCATCTTTGCC 251
QY 255 ACCAAGCCTCATGTGCTGCGCAGCTGCTCTGCGCAGAGTGGCTCTGTCACACTGAA 314
DB 252 ACCAAGCCTCATGTGCTGCGCAGCTGCTCTGCGCAGAGTGGCTCTGTCACACTGAA 311
QY 315 CACATCTTGGTGTCCGTGCTGCTGGGTGTCTGTGAGCACCCTGGAGAGTGTGGCC 374
DB 312 CACATCTTGGTGTCCGTGCTGCTGGGTGTCTGTGAGCACCCTGGAGAGTGTGGCC 371
QY 375 CCAACACACGCGGTGTGGGGTCTTGCCCAACCTGCTGTGAGTGTGGTCCAGAAAGGGCC 434
DB 372 CCAACACACGCGGTGTGGGGTCTTGCCCAACCTGCTGTGAGTGTGGTCCAGAAAGGGCC 431
QY 435 ATAGTGTATGGCGGGCGCGCCACCTGGGAGCAGCAGACCAAGCTCC-TGCAGCATCT 493
DB 432 ATAGTGTATGGCGGGCGCGCCACCTGGGAGCAGCAGACCAAGCTCCCTTTGCGAGCATCT 491
QY 494 GCTGAGGCGCTGTGGCGGTGTGAGGAGTGTGCTGAAGCCTACGTCGACATCCACTGG 553
DB 492 GCTGAGGCGCTGTGGCGGTGTGAGGAGTGTGCTGAAGCCTACGTCGACATCCACTGG 551
QY 554 CTTAGTGGCAGTGGCTGGCTTTCGTGTGTCATTCCTGAGGCGCTTGGCTGAAGGAGC 613
DB 552 CTTAGTGGCAGTGGCTGGCTTTCGTGTGTCATTCCTGAGGCGCTTGGCTGAAGGAGC 611
QY 614 CTTAGTGGCAGTGGCTGGCTTTCGTGTGTCATTCCTGAGGCGCTTGGCTGAAGGAGC 673
DB 612 CTTAGTGGCAGTGGCTGGCTTTCGTGTGTCATTCCTGAGGCGCTTGGCTGAAGGAGC 671
QY 674 GACGCCCAAGATGCTGCTGCACGAGGGCCAAACCCAGCCAGCTGGGCTCAGACGTGTG 733
|||||

DB 672 GACGCCCAAGATGCTGCTGCACGAGGGCCAAACACCAGCCAGCTGNGCTCAGACGTGTG 731
QY 734 CA-CCCCGGGTGGCACCACCATCTATGGACTTCAGCGCCCTGGACAGCGGGGGCTGGCAG 792
DB 732 CACCCCGGGTGGGAGCCACCATCTATGGACTTCAGCGCCCTGGACAGCGGGGGCTGNN 791
QY 793 CAG 795
DB 792 CAG 794
RESULT 10
BM473687
LOCUS 1145 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6466593 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5562066
5', mRNA sequence.
ACCESSION BM473687
VERSION BM473687.1 GI:18522729
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1145)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM12290 row: n column: 19
High quality sequence stop: 503.
Location/Qualifiers
1. .1145
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5562066"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 Kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 194 a 392 c 356 g 195 t 8 others
ORIGIN
Query Match 52.4%; Score 701.2; DB 13; Length 1145;
Best Local Similarity 96.5%; Pred. No. 1.1e-147;
Matches 727; Conservative 0; Mismatches 24; Indels 2; Gaps 1;
QY 24 AACAAATGTCAGCTCGGAGCGGTCTCCGCGCGGTGGCTTCGTGGCGGGGGCCG 83
DB 1 AACAAATGTCAGCTCGGAGCGGTCTCCGCGCGGTGGCTTCGTGGCGGGGGCCG 60
QY 84 ATGCGGGGGCCATCGCGAGGCGCTCATCAGACGAGAAAGTGAAGCTCAGCACATA 143
DB 61 ATGCGGGGGCCATCGCGAGGCGCTCATCAGACGAGAAAGTGAAGCTCAGCACATA 120
QY 144 CTGCGCAGTGCACCAACACAGACAGGAACTATGTCATTTCAAGCTCTGGGTGCGGAGC 203
DB 121 CTGCGCAGTGCACCAACACAGACAGGAACTATGTCATTTCAAGCTCTGGGTGCGGAGC 180
QY 204 ACACACTTCAACACAGGAGGTGTGTCAGAGTGCCTGCTCATCTTTGCCACCAAGCCT 263
DB 181 ACACACTTCAACACAGGAGGTGTGTCAGAGTGCCTGCTCATCTTTGCCACCAAGCCT 240

264 CATGTGCTGCCAGCTCTCTCTGGAGAGTGGCTCTGTGGTCCACCACTGAACACATCTTGG 323
|||||
241 CATGTGCTGCCAGCTCTCTCTGGAGAGTGGCTCTGTGGTCCACCACTGAACACATCTTGG 300
|||||
324 GTTCCGCTGGCTCTCTGGGTGTCTCTGAGCACCTGGAGAGTGTGTGCCCCCAACACA 383
|||||
301 GTTCCGCTGGCTCTCTGGGTGTCTCTGAGCACCTGGAGAGTGTGTGCCCCCAACACA 360
|||||
384 CGGGTCTGGGGTCTTGGCCCAACCTGCCCCCTGTGTGGTCCAGAAAGGGGCCATAGTGATG 443
|||||
361 CGGGTCTGGGGTCTTGGCCCAACCTGCCCCCTGTGTGGTCCAGAAAGGGGCCATAGTGATG 420
|||||
444 GCSCGGGGCGCCACCTGGGGAGCAGCAGAGCAAGCTCCTGCAGCATCTGTGGAGGCC 503
|||||
421 GCSCGGGGCGCCACCTGGGGAGCAGCAGAGCAAGCTCCTGCAGCATCTGTGGAGGCC 480
|||||
504 TGTGGGCGGTGTGAGGAGTGGCTGAAGCTTACGTGAGCATCCACACTGGCCCTCAAGATG 523
|||||
541 AGTGGGCGGTGGCTTCTGATTTCTCGAGGCCCTGGCTGAAGGAGCCGCTCAAGATG 600
|||||
624 GGCATGCCAGCAGCTGGCCACCGCATCGCTGCCCCAGACCTGCTGGGGAGCGGCAAG 683
|||||
601 GGCATGCCAGCAGCTGGCCACCGCATCGCTGCCCCAGACCTGCTGGGGAGCGGCAAG 660
|||||
684 ATGCTGCTGACAGAGGGCCAAACACCCAGCCAGCTGCGCTGAGACGTGTGCACCCC - -GG 741
|||||
661 ATGCTGCTGACAGAGGGCCAAACACCCAGCCAGCTGCGCTCAAACTGGGCCCCCCCGG 720
|||||
742 GTGGCACCACCATCTATGACTCCAGCGCTGG 774
|||||
721 GGGGAGCCCATCTATGACTCCAGCGCTTG 753
|||||

RESULT 11
BQ436653 932 bp mRNA linear EST 24-MAY-2002
LOCUS AGENCOURT_7801851 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6050678
5', mRNA sequence.
ACCESSION BQ436653
VERSION BQ436653.1 GI:21175729
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 932)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/ntp

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLAM13303 row: e column: 15
High quality sequence stop: 565.

FEATURES
source

Location/Qualifiers
1..932
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6050678"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"

/note="organ: skin; Vector: pCMV-SPORT6; Site:1: Noti;
Site_2: SalI; Cloned unidirectionally. primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT	163 a	311 c	306 g	152 t
ORIGIN				
Query Match		52.1%	Score 697.4;	DB 14; Length 932;
Best Local Similarity		97.0%;	Pred: No. 7e-147;	
Matches 732; Conservative		0;	Mismatches 21;	Indels 2; Gaps 2;
QY	41	GGAGCGCTCTCGGGCGCGTGGGCTTCTGTGGGGCGGGCGGCATGGCGGGGGCCATCGC	100	
Db	1	GGAGCGCTCTCGGGCGCGTGGGCTTCTGTGGGGCGGGCGGCATGGCGGGGGCCATCGC	60	
QY	101	CGAGGGCTCATCAGAGCAGGAGAAAGTGAAGTCTAGACATCTACTGGCCATGTCCACCAAC	160	
Db	61	CGAGGGCTCATCAGAGCAGGAGAAAGTGAAGTCTAGACATCTACTGGCCATGTCCACCAAC	120	
QY	161	AGACAGGAACCTATGTCACTTTCAAGCTCTGGGTGCGGACACGACACTCCAAACACAGA	220	
Db	121	AGACAGGAACCTATGTCACTTTCAAGCTCTGGGTGCGGACACGACACTCCAAACACAGA	180	
QY	221	GGTGTCTGAGAGCTGCTCTGCTGCTATCTTTGCCAACCAAGCTCATGTGTGCCAGCTGT	280	
Db	181	GGTGTCTGAGAGCTGCTCTGCTGCTATCTTTGCCAACCAAGCTCATGTGTGCCAGCTGT	240	
QY	281	CCTGCAGAGTGGCTCTCTGCTGCTACCACTGAACACATCTTGGTGTCCGTGGCTGGTGG	340	
Db	241	CCTGCAGAGTGGCTCTCTGCTGCTACCACTGAACACATCTTGGTGTCCGTGGCTGGTGG	300	
QY	341	GGTGTCTCTGAGCACCCTGGAGAGCTGCTGCCCCCAACACACAGGGTGTGCGGGTCTT	400	
Db	301	GATGTCTCTGAGCACCCTGGAGAGCTGCTGCCCCCAACACACAGGGTGTGCGGGTCTT	360	
QY	401	GCCCAACCTGCCCTGTGTGGTCCAGAAAGGGGCCATAGTGTGCGGGGGCGCCACGT	460	
Db	361	GCCCAACCTGCCCTGTGTGGTCCAGAAAGGGGCCATAGTGTGCGGGGGCGCCACGT	420	
QY	461	GGGAGCAGCAGACCAAGCTCTCTGCAGCATCTGTGGAGGCTGTGGGGCGGTGTGAGGA	520	
Db	421	GGGAGCAGCAGACCAAGCTCTCTGCAGCATCTGTGGAGGCTGTGGGGCGGTGTGAGGA	480	
QY	521	GGTGTCTGAAGCTACGTCCACATCCACATCGGCTCAGTGGCAGTGGCTGGCTTCGT	580	
Db	481	GGTGTCTGAAGCTACGTCCACATCCACATCGGCTCAGTGGCAGTGGCTGGCTTCGT	540	
QY	581	GTGTGCATTTCTCCGAGGCCCTGGCTGAAGAGGCGCTCAAGATGGCATGCCACAGCGCT	640	
Db	541	GTGTGCATTTCTCCGAGGCCCTGGCTGAAGAGGCGCTCAAGATGGCATGCCACAGCGCT	600	
QY	641	GGCCACCGCATCTGCCAGACCTGCT-GGGACGCGCAAGATGCTGTGCACAGG	699	
Db	601	GGCCACCGCATCTGCCAGACCTGCT-GGGACGCGCGCAAGATGCTGTGCACAGG	660	
QY	700	GGCAACACCCAGCCAGCTGGCTCAGACGTGTGCACCCC-GGGTGGCACCCATCTAT	758	
Db	661	GGCAACACCCAGCCAGCTGGCTCAGACGTGTGCACCCC-GGGTGGCACCCATCTAT	720	
QY	759	GGACTCCAGCCCTGGAGCGGGCGGTGGCAGC	793	
Db	721	GGACTCCAGCCCTGGGAAACCGAGGGGGGCTGC	755	

RESULT 12
BI253559
LOCUS 602973485F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5112692 5',
DEFINITION mRNA sequence.
ACCESSION BI253559
VERSION BI253559.1 GI:14805103
KEYWORDS EST.
SOURCE human.

ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .									
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)									
JOURNAL	Unpublished (1999)									
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Incyte Genomics, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM11273 row: j column: 21 High quality sequence stop: 734.									
FEATURES	Location/Qualifiers									
source	1..734									
	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
	/clone="IMAGE:5112692"									
	/clone_lib="NIH_MGC_12"									
	/tissue_type="cervical carcinoma cell line"									
	/lab_host="DH10B"									
	/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.4 kb. Library prepared by Life Technologies."									
BASE COUNT	127	a	228	c	253	g	126	t		
ORIGIN										
Query Match	51.4%; Score 688.2; DB 13; Length 734;									
Best Local Similarity	98.6%; Pred. No. 7.4e-145;									
Matches	715;	Conservative	0;	Mismatches	8;	Indels	2;	Gaps	2;	
QY	36	GCTGGGAGCGCTCTCCGGCGCGTGGGCTTCGTGGCGGGCGCGCATGGCGGGGCC	95							
Db	11	GGTGGGAGCCGCTCTCCGGCGCGGTGGGGTCTGTGGCGGGCGCGCATGGCGGGGCC	70							
QY	96	ATCGGCGAGGGCCTCATCAGACGAGGAAAAGTGAAGCTCAGCACATACCTGGCGAGTGCA	155							
Db	71	ATCGCGCAGGGCCTCATCAGACGAGGAAAAGTGAAGCTCAGCACATACCTGGCGAGTGCA	130							
QY	156	CCAACAGACAGGAACCTATGTACATTTCAAGCTCTGGGTTGCCGACACAGCACTCCAAC	215							
Db	131	CCAACAGACAGGAACCTATGTACATTTCAAGCTCTGGGTTGCCGACACAGCACTCCAAC	190							
QY	216	CAGGAGGTGCTGCAGAGCTGCCTGCTCGTATCTTTGCCACCAAGCCTATGTGCTGCCA	275							
Db	191	CAGGAGGTGCTGCAGAGCTGCCTGCTCGTATCTTTGCCACCAAGCCTATGTGCTGCCA	250							
QY	276	GCTGTCTCGCAGAGTGGCTCTGTGGTCAACACTGAACACATCTTTGGTGTCCGTGGCT	335							
Db	251	GCTGTCTCGCAGAGTGGCTCTGTGGTCAACACTGAACACATCTTTGGTGTCCGTGGCT	310							
QY	336	GCTGGGTGTCTCTGAGCACCCCTCGAGAGCTGCTGCCCCCAACACACAGGGTGTCTGGC	395							
Db	311	GCTGGGTGTCTCTGAGCACCCCTCGAGAGCTGCTGCCCCCAACACACAGGGTGTCTGGC	370							
QY	396	GTCTTCCCAACCTCCCTCTGTGTGTCAGAGGGGCCATAGTATGGCGGGGCCG	455							
Db	371	GTCTTCCCAACCTCCCTCTGTGTGTCAGAGGGGCCATAGTATGGCGGGGCCG	430							
QY	456	CACGTGGGAGCAGCAGACCAAGCTCCTCGAGCATCTGCTGAGGCGCTGTGGCGGTGT	515							
Db	431	CACGTGGGAGCAGCAGACCAAGCTCCTCGAGCATCTGCTGAGGCGCTGTGGCGGTGT	490							
QY	516	GAGGAGGTCCCTGAAGCCTACGTCCGATCCCACTGGCCTCACTGGCAGTGGCGTGGCC	575							
Db	491	GAGGAGGTCCCTGAAGCCTACGTCCGATCCCACTGGCCTCACTGGCAGTGGCGTGGCC	550							

QY	576	TTCGTGTTGCATTCTCCGAGGCCCTGGCTCAAGAGAGCCGTC	AAGATGGGCATGCCCAGC	633
Dd	551	TTCGTGTTGCATTCTCCGAGGCCCTGGCTCAAGAGAGCCGTC	AAGATGGGCATGCCCAGC	610
QY	636	AGCCTGGCCCCACCGCATCGCTGCCAGACACCCTGCTGGGAGCGGCCAAGATGCTGCTGCAC	695	
Dd	611	AGCCTGGCCCCACCGCATCGCTGCCAGACACCCTGCTGGGAGCGGCCAAGATGCTGCTGCAC	669	
QY	696	G-AGGGCCAACACCCAGCCAGCTGCGCTCAGACTGTGTGCACCCCGGGTGGCACACCAT	754	
Dd	670	GAAGGGCCAAACCCAGCCAGCTGCGCTCAGACTGTGTGCACCCCGGGTGGCACACCAT	729	
QY	755	CTATG 759		
Dd	730	CTATG 734		
RESULT 13				
AL522085				
LOCUS				
DEFINITION	AL522085 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB007YC12 5 prime, mRNA sequence.	697 bp	mRNA linear	EST 13-FEB-2001
ACCESSION	AL522085			
VERSION	AL522085.1 GI:12785578			
TITLE	EST.			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	L.I.W.B., Gruber,C., Jessee,J. and Polayes,D.			
JOURNAL	Full-length cDNA libraries and normalization Unpublished (2001)			
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr. Location/Qualifiers			
FEATURES	1..697			
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="CS0DB007YC12" /clone_lib="LTI_NFL004_NBC2" /sex="male" /tissue_type="neuroblastoma cells" /lab_host="DH10B"			
	/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"			
BASE COUNT	123 a 219 c 233 g 121 t			1 others
ORIGIN				
Query Match	50.3%	Score 673.6;	DB 9;	Length 697;
Best Local Similarity	99.6%	Pred. No. 1.4e-141;		
Matches 695;	Conservative	1;	Mismatches 0;	Indels 2;
Gaps	2;			
QY	23	CACAAGATGGCAGCTCGGAGCGGTCTCGGCGCGGTGGGCTTCGTGGCGGGCGCG	82	
Dd	1	CAACAAGATGGCAGCTCGGA-CGGTCTCGGCGCGGTGGGCTTCGTGGCGGGCGCG	59	
QY	83	CATGGCGGGGCCATCGCGAGGCGCTCATCAGACAGGAAAAGTGAAGCTCAGACAT	142	
Dd	60	CATGGCGGGGCCATCGCGAGGCGCTCATCAGACAGGAAAAGTGAAGCTCAGACAT	119	
QY	143	ACTGGCCAGTGCCACCAACAGACAGAACCTATGTCACTTTCAAAGCTCTGGTTGCCGAC	202	

Db	120	ÄTGGCCAGTGCACCAACAGACAGAAACCTATGTCACTTTCAAGCTCTGGGTTGCGGAC	179
Qy	203	CACGCACCTCAACACGAGAGGTGCTGCAGAGTGCCTGCTCGTCTCATCTTTGGCCACCAAGCC	262
Db	180	CAGGCACCTCAACACGAGAGGTGCTGCAGAGTGCCTGCTCGTCTCATCTTTGCCACCAAGCC	239
Qy	263	TCATGTGCTGCCAGCTGCTCCTGGCAGAGGTGGCTCTGTGGTCACCACCTGAACACATCTTT	322
Db	240	TCATGTGCTGCCAGCTGCTCCTGGCAGAGGTGGCTCTGTGGTCACCACCTGAACACATCTTT	299
Qy	323	GGTGTCCGTGGCTGCTGGGGTGTCTCTGAGCACCCCTGGAGGAGCTGCTGCCCCCAACAC	382
Db	300	GGTGTCCGTGGCTGCTGGGGTGTCTCTGAGCACCCCTGGAGGAGCTGCTGCCCCCAACAC	359
Qy	383	ACGGGTGCTCGGGGTCTTGCCCAACTGCCCTGTGTGTCAGGAAGGGGCCATAGTGAAT	442
Db	360	ACGGGTGCTCGGGGTCTTGCCCAACTGTGCTGTGTGTCAGGAAGGGGCCATAGTGAAT	419
Qy	443	GGCGGGGGGC - GCCCACCTGGGGAGCAGCAGACCAAGCTCTCTGCAGCATCTGCTGGAGG	501
Db	420	GGCGGGGGCGGCCACCTGGGGAGCAGCAGACCAAGCTCTCTGCAGCATCTGCTGGAGG	479
Qy	502	CCTGTGGGCGGTGTGAGGAGTGGCTGAAGCCTACGTCTGACATCCACACTGGGCTCAGTG	561
Db	480	CCTGTGGGCGGTGTGAGGAGTGGCTGAAGCCTACGTCTGACATCCACACTGGGCTCAGTG	539
Qy	562	GCATGTGGCTGCGCCTTCGTGTGTCATCTCCGAGGGCCCTGGCTGAAGGAGCGGTCAAGA	621
Db	540	GCATGTGGCTGCGCCTTCGTGTGTCATCTCCGAGGGCCCTGGCTGAAGGAGCGGTCAAGA	599
Qy	622	TGGGCATGCCAGCAGCTGGCCACCGCATCGTGCCTGCCAGACCCCTGCTGGGGACGGCCA	681
Db	600	TGGGCATGCCAGCAGCTGGCCACCGCATCGTGCCTGCCAGACCCCTGCTGGGGACGGCCA	659
Qy	682	AGATGCTGCTGCACGAGGGCCAAACACCCAGCCCAAGCTG	719
Db	660	AGATGCTGCTGCACGAGGGCCAAACACCCAGCCCAAGCTG	697

BM010281 741 bp mRNA linear EST 30-OCT-2001
603631154F1 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5444893 5',
mRNA sequence.

BM010281
BM010281.1 GI:16524635
EST.

human.

SOURCE
ORGANISM

homo sapiens
Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 741)

I (bases 1 to 41)
NTH-MGC <http://mgc.ncbi.nih.gov/>

NTH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health Mammalian Gene Collection (MGC)

National Institutes of Health, M...

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTP

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arranged by: The T M A C F
CDNA Library Preparation: Liny HONG/KNUST

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution:

found through the I.M.A.G.E. Consor

<http://image.llnl.gov>

Plate: LLCM1924 row: 1 col: 1

High quality sea

FEATURES

FIGURES

1. 741

```
/organism="Homo sapiens"
```

/db_xref="taxon:9606"

```
/clone="IMAGE:5444893"
```

```
/ clone lib="NTH MGC 41"
```

RESULT 15

RESULTS
PT 8198

5818TB

LOCUS

DEFINITION 603033078F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5174178 5',

mRNA sequence.
BI818555
VERSION BI818555.1 GI:15929661
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 773)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11433 row: 1 column: 19
High quality sequence stop: 763.
Location/Qualifiers
1..773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5174178"
/clone_lib="NIH_MGC_115"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."
146 a 237 c 257 g 133 t

FEATURES

Source

BASE COUNT
ORIGIN

Query Match 46.9%; Score 627.6; DB 13; Length 773;
Best Local Similarity 96.2%; Pred.No. 3.6e-131;
Matches 729; Conservative 0; Mismatches 19; Indels 10; Gaps 8;
QY 15 GTCCGAGGCAACAGATGGCAGCTGCGGAGCGCTCTCCGGCGCGCTGGGCTTCGTGGGC 74
DB 12 GTCCGAGGCAACAGATGGCAGCTGCGGAGCGCTCTCCGGCGCGCTGGGCTTCGTGGGC 71
QY 75 GCGGCGCGCATGGCGGGGCCATCGCGCAGGGCCCTCATCAGACGAGGAAAGTGGAAAGCT 134
DB 72 GCGGCGCGCATGGCGGGGCCATCGCGCA-GGCCTCATCAGACGAGGAAAGTGGAAAGCT 130
QY 135 CAGCACATCTGCGCAGTGCCACACACAGACAGAGAACCTATGTCACCTTCAAGCTCTGGGT 194
DB 131 CAGCACATCTGCGCAGTGCCACACACAGACAGAGAACCTATGTCACCTTCAAGCTCTGGGT 190
QY 195 TGCCGGACAGGCACTCAACACAGGAGGTGCTGCAGAGCTGCCCTGCTCGTCACTTTTGCC 254
DB 191 TGCCGGACAGGCACTCAACACAGGAGGTGCTGCAGAGCTGCCCTGCTCGTCACTTTTGCC 250
QY 255 ACCAAGCCTCATGTGCTGCCAGCTGTCTTGGCAGAGGTGGCTCCTGTGGTCAACACTGAA 314
DB 251 ACCAAGCCTCATGTGCTGCCAGCTGTCTTGGCAGAGGTGGCTCCTGTGGTCAACACTGAA 310
QY 315 CACATCTTGGTGTCGCTGGCTGCTGGGTGCTCTGTGACAGAGGTGGCTCCTGTGGTCAACACTGAA 374
DB 311 CACATCTTGGTGTCGCTGGCTGCTGGGTGCTCTGTGACAGAGGTGGCTCCTGTGGTCAACACTGAA 370
QY 375 CCAACACACGGGTGCTCGGGTCTTGCCCAACCTGCCCTGTGTGGTCCAGGAAGGGGCC 434

Db 371 CCAACACACGGGTGCTCGGGTCTTGCCCAACCTGCCCTGTGTGGTCCAGGAAGGGGCC 430
QY 435 ATAGTGATGGCGGGGCCGCCACAGTGGGAGGAGAGACCAAGCTCCTGTCAGCA--TC 492
Db 431 ATAGTGATGGCGGGGCCG-CACGTGGGAGGAGAGACCAAGCTCCTGTCAGCATCTC 489
QY 493 TGTGGAGGCTGTGGCGGTGTGAGGAGTGGCTTGAAGCTTACGTGCGACATCCACACTG 552
Db 490 TGTGGAGGCTGTGGCGGTGTGAGGAGTGGCTTGAAGCTTACGTGCGACATCCACACTG 549
QY 553 GCCTCAGTGGCAG-TGGCGTGGCTTGGTGTGTGCAATTCCTCCAGGCGCTGGCTGAAGGA 611
Db 550 GCCTCAGTGGCAGTGGCGTGGCTTGGTGTGTGCAATTCCTCCAGTGGCTGAAGGA 609
QY 612 GCCGTCAAGATGGCATGCCACAGCCTGG-CCCACCGCATGCTGCCAGACCCCTGCT 670
Db 610 GCCGTCAAGATGGCATGCCACAGCCTGGACCTGACAGATCGCTGCCAGAGCCTGCT 669
QY 671 GGGGA-CGGCCAAAGATGCTGCTGCAC--GAGGGCCAAACACCCAGCCAGCTGCGCTCAG- 726
Db 670 GGGGACCGGCAAGATGCTGCTGCACAGAGGGGCCAAGAACCCAGGCCAAGCTGCGCTCAGA 729
QY 727 ACGTGTGCACCCCGGGTGGCACCACCATCTATGGACTC 764
Db 730 ACGTGTGCACCCCGGGTGGCACCACCATCTATGGACTC 767

Search completed: July 26, 2003, 07:20:42
Job time : 2127 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: July 21, 2003, 09:48:21 ; Search time 80 Seconds
(without alignments)
705.711 Million cell updates/sec

Title: US-09-806-536A-14
Perfect score: 1386
Sequence: 1 MAABEPSRRVGVGAGRMA.....AATMSAVEAATCRAKELSRK 274

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1386	100.0	274	4	Q9H896
2	1381	99.6	274	4	Q96HX4
3	1163	83.9	274	11	Q9DCC4
4	1160	83.7	274	11	Q8R0P9
5	921	66.5	248	11	Q9D0X2
6	613	44.2	279	5	Q21544
7	577.5	41.7	320	4	Q96C36
8	571.5	41.2	320	11	Q922Q4
9	560.5	40.4	273	5	Q9VEJ3
10	560.5	40.4	309	11	Q922W5
11	533.5	38.5	270	17	Q9HH99
12	505.5	36.5	280	5	Q9V3F8
13	503	36.3	274	10	Q9AYM3
14	493.5	35.6	280	5	Q96B43
15	480	34.6	270	16	Q97E64
16	477	34.4	319	4	Q9Y5J4

ALIGNMENTS

RESULT 1

Q9H896 PRELIMINARY; PRT; 274 AA.
AC Q9H896;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THYR01000934 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=THYROID;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NEO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK023914; BAB14721.1;
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR PROSITE; PS00521; P5CR; 1.
SQ SEQUENCE 274 AA; 28663 MW; 846FDEC603F3B548 CRC64;

Query Match 100.0%; Score 1386; DB 4; Length 274;

Best Local Similarity 100.0%; Pred. No. 1.4e-100; Indels 0; Gaps 0;

Matches 274; Conservative 0; Mismatches 0;

QY 1 MAABEPSRRVGVGAGRMAIAOGLIRAGKVEAQHILASAPTDRNLCHFQALGCRTH 60

|||||

Db 1 MAABEPSRRVGVGAGRMAIAOGLIRAGKVEAQHILASAPTDRNLCHFQALGCRTH 60

|||||

QY 61 SNOEVLQSCLLVIFATKPHVLPVLAIEAVPVVTTTHILVSAAGVSLSTLELLPPNTRV 120

|||||

Db 61 SNOEVLQSCLLVIFATKPHVLPVLAIEAVPVVTTTHILVSAAGVSLSTLELLPPNTRV 120

|||||

QY 121 LRVLPNLCVVOEGAIVMARGHVGSSSETKLLQHLLEACGRCEEVPEAYVDIHTGLSGS 180
Db 121 LRVLPNLCVVOEGAIVMARGHVGSSSETKLLQHLLEACGRCEEVPEAYVDIHTGLSGS 180
QY 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
Db 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
QY 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274
Db 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274

RESULT 2

Q96HX4 ID Q96HX4 PRELIMINARY; PRT; 274 AA.
AC Q96HX4;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical 28.6 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007993; AAH07993.1; -
DR InterPro: IPR000304; P5CR.
DR Pfam: PF01089; P5CR; 1.
DR PROSITE: PS00521; P5CR; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 274 AA; 28649 MW; 846FD9B60183B048 CRC64;

Query Match 99.6%; Score 1381; DB 4; Length 274;
Best Local Similarity 99.6%; Pred. No. 3.5e-100;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MAAEPPSRVGVGAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNLCFQALGCRTH 60
Db 1 MAAEPPSRVGVGAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNLCFQALGCRTH 60
QY 61 SNQEVLOSCLLVIFATKPHVLPVLAEPVVTTEHILVSVAAAGVSLTLEELPPNTRY 120
Db 61 SNQEVLOSCLLVIFATKPHVLPVLAEPVVTTEHILVSVAAAGVSLTLEELPPNTRY 120
QY 121 LRVLPNLCVVOEGAIVMARGHVGSSSETKLLQHLLEACGRCEEVPEAYVDIHTGLSGS 180
Db 121 LRVLPNLCVVOEGAIVMARGHVGSSSETKLLQHLLEACGRCEEVPEAYVDIHTGLSGS 180
QY 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
Db 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
QY 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274
Db 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274

RESULT 3

Q9DCC4 ID Q9DCC4 PRELIMINARY; PRT; 274 AA.
AC Q9DCC4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 1110058B13rik protein.
GN 1110058B13rik.
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
EX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bernaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK002912; BAB22451.1; -
DR MGD; MGI:1913444; 1110058B13rik.
DR InterPro: IPR000304; P5CR.
DR Pfam: PF01089; P5CR; 1.
SQ SEQUENCE 274 AA; 28694 MW; FE3892C01C6068A5 CRC64;
Query Match 83.9%; Score 1163; DB 11; Length 274;
Best Local Similarity 82.5%; Pred. No. 3.7e-83;
Matches 226; Conservative 23; Mismatches 25; Indels 0; Gaps 0;
QY 1 MAAEPPSRVGVGAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNLCFQALGCRTH 60
Db 1 MAATMSEPRVGVGAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNLCFQALGCRTH 60
QY 61 SNQEVLOSCLLVIFATKPHVLPVLAEPVVTTEHILVSVAAAGVSLTLEELPPNTRY 120
Db 61 SNQEVLOSCLLVIFATKPHVLPVLAEPVVTTEHILVSVAAAGVSLTLEELPPNTRY 120
QY 121 LRVLPNLCVVOEGAIVMARGHVGSSSETKLLQHLLEACGRCEEVPEAYVDIHTGLSGS 180
Db 121 LRVLPNLCVVOEGAIVMARGHVGSSSETKLLQHLLEACGRCEEVPEAYVDIHTGLSGS 180
QY 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
Db 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
QY 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274
Db 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274
RESULT 4
Q8R0P9 ID Q8R0P9 PRELIMINARY; PRT; 274 AA.
AC Q8R0P9;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE RIKEN cDNA 1110058B13 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;

```
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC026536; AAH26536.1; -.
SQ SEQUENCE 274 AA; 28721 MW; E01FB7133B45BD7C CRC64;

Query Match      83.7%; Score 1160; DB 11; Length 274;
Best Local Similarity 82.1%; Pred. No. 6.3e-83;
Matches 225; Conservative 24; Mismatches 25; Indels 0; Gaps 0;

QY 1 MAAAEPSRRVGVGAGRMAGAIAOGLIRAGKVEAQHILASAPTDRNLCHFQALGCRTH 60
DB 1 MAATMSPRRVGVGAGRMAGAEIARGLIQAGKVEAQVLAASAPTDNNLCHFALGCRTH 60
QY 61 SNOEVLQSCLLVIFATKPHVLPVLAEPVVTTEHILVSAAGVSLSTLEELLPPNTRV 120
DB 61 SNEVLQNCPLVIFATKPHVLPVLAEPVVTTEHILVSAAGVSLSTLEELLPPNTRV 120
QY 121 LRVLPNLCVVOEGAIVMARGHVSGSETKLLQHLLEACRCCEEVPEAYVDIHTGLSGSG 180
DB 121 LRVLPNLCVVOEGAIVMARGHVSGSETKLLQHLLEACRCCEEVPEAYVDIHTGLSGSG 180
QY 181 VAFVCAFESEALAEAGVKGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
DB 181 VAFVCTFESEALAEAGVKGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
QY 241 TIYGLHAL 248
DB 237 TCWNHSM 244

RESULT 6
Q21544 PRELIMINARY; PRT; 279 AA.
AC Q21544;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE M153.1 protein.
GN M153.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews P.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z67995; CAA91943.1; -.
DR InterPro: IPR000304; P5CR.
DR Pfam: PF01089; P5CR; 1.
DR TIGRFAMs: TIGR00112; proc; 1.
SQ SEQUENCE 279 AA; 29165 MW; C8C8AD4BC4EB20E6 CRC64;

Query Match      44.2%; Score 613; DB 5; Length 279;
Best Local Similarity 46.1%; Pred. No. 3.3e-40;
Matches 123; Conservative 56; Mismatches 86; Indels 2; Gaps 1;

QY 10 RVGVGAGRMAGAIAOGLIRAGKVEAQHILASAP--TDRLNCHFQALGCRTHSNOEVLQ 67
DB 2 KICFIGAGKMAQALAGLINSGRITADNIASSPKRDEFLDQCKALGLTTHDIAEVVQ 61
QY 68 SCILVIFATKPHVLPVLAEPVVTTEHILVSAAGVSLSTLEELLPPNTRVLPN 127
DB 62 KSDVFLAVKPHVSVKVAISEIAPLSALGIRIESLLPTIRIESLLPTIRIESLLPTIR 121
QY 128 PCVVOEGAIVMARGHVSGSETKLLQHLLEACRCCEEVPEAYVDIHTGLSGSGVAFVCAF 187
DB 122 PSVVRAGAFAMGSCNCRDGAETVEKLLSTYGVFAVEPEIHDVPTGLSGSGSYMFAV 181
QY 188 SEALAEAGVKGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 247
DB 188 SEALAEAGVKGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 247

RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC026536; AAH26536.1; -.
SQ SEQUENCE 274 AA; 28721 MW; E01FB7133B45BD7C CRC64;

Query Match      83.7%; Score 1160; DB 11; Length 274;
Best Local Similarity 82.1%; Pred. No. 6.3e-83;
Matches 225; Conservative 24; Mismatches 25; Indels 0; Gaps 0;

QY 1 MAAAEPSRRVGVGAGRMAGAIAOGLIRAGKVEAQHILASAPTDRNLCHFQALGCRTH 60
DB 1 MAATMSPRRVGVGAGRMAGAEIARGLIQAGKVEAQVLAASAPTDNNLCHFALGCRTH 60
QY 61 SNOEVLQSCLLVIFATKPHVLPVLAEPVVTTEHILVSAAGVSLSTLEELLPPNTRV 120
DB 61 SNEVLQNCPLVIFATKPHVLPVLAEPVVTTEHILVSAAGVSLSTLEELLPPNTRV 120
QY 121 LRVLPNLCVVOEGAIVMARGHVSGSETKLLQHLLEACRCCEEVPEAYVDIHTGLSGSG 180
DB 121 LRVLPNLCVVOEGAIVMARGHVSGSETKLLQHLLEACRCCEEVPEAYVDIHTGLSGSG 180
QY 181 VAFVCAFESEALAEAGVKGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
DB 181 VAFVCTFESEALAEAGVKGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
QY 241 TIYGLHAL 248
DB 237 TCWNHSM 244

RESULT 5
Q9D0X2 PRELIMINARY; PRT; 248 AA.
AC Q9D0X2;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE I110058B13R1k protein.
GN I110058B13R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuenli P., Lewis S., Matsuo Y., Nikolaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momabaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wvshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK004291; BAB23252.1; -.
DR MGD; MGI:1913444; I110058B13R1k.
DR InterPro: IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
SQ SEQUENCE 248 AA; 26170 MW; 4AFA896396305A34 CRC64;
```

```
Db 182 IEGLADGVKVGVLPRDLALKLAAAYTLGLGAAMVLEGTIIHPAQLKDDVQSPAGSSVYGMHK 241
QY 248 LEQGLRAATMSAVEAATCRAKELSRK 274
DB 242 LESGGLKGYLMDAVEAATNRSRATGDK 268

RESULT 7
Q96C36 PRELIMINARY; PRT; 320 AA.
AC Q96C36;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Similar to pyroline 5-carboxylate reductase isoform (Hypothetical
DE 33.6 kDa protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE=CERVIX;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014868; AAH14868.1; -.
DR EMBL; BC020553; AAH20553.1; -.
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR PROSITE; PS00521; P5CR; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 320 AA; 33637 MW; 9348455A6ACCE722 CRC64;

Query Match 41.7%; Score 577.5; DB 4; Length 320;
Best Local Similarity 46.4%; Pred. No. 2.2e-37;
Matches 124; Conservative 48; Mismatches 88; Indels 7; Gaps 3;

QY 11 VGFVAGRMAGATAGLIRAGKVEAQHILASAPTDRNLCHFQAL---GCRTHSNQEVQLQ 67
DB 3 VGFAGQAYALAGFTAAAGLSAHKIIASSP-EMNLPTVSALRMGMVNLTRSNKETYK 61
QY 68 SCLLVIFATKPHVLPVLAELAVPVVTEHILYSVAAGVSLTLEELL---PNTNRVLRVL 124
DB 62 HSDVLFVAVKPHIIPFILDIGADVQARHIVVSCAAGVTISSVEKKLMAFQAPKVICM 121
QY 125 PNLPCVQBGAIVMARGHRVGSSETKLQHLLEACRCCEVPEAYVDIHTGLSGGVAFV 184
DB 122 TNPVVRGATVYATGTHALVEDGKLLBQLMSSVGFCTEVEDLIDAVTGLSGGPAYA 181
QY 185 CAFSEALARGAVKMGMPSSLAHRIAATLLGTAKMLLHGGHQAQLRSVDCVTPGGTTIYG 244
DB 182 FMALDALADGVKMGVPRRLAQLGAALIGAALMLLSDHDPGQLKDNVCSPPGGATIHA 241
QY 245 LHALEGGGLRAATMSAVEAATCRAKEL 271
DB 242 LHFLSGGFRSLINAVEASCIRTREL 268

RESULT 8
Q922Q4 PRELIMINARY; PRT; 320 AA.
AC Q922Q4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to pyroline 5-carboxylate reductase isoform.
GN 1810018M05RIK.
OS Mus musculus (Mouse).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN 1
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006882; AAH06882.1; -.
DR MGD; MGI:1916301; 1810018M05RIK.
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR PROSITE; PS00521; P5CR; UNKNOWN_1.
SQ SEQUENCE 320 AA; 33659 MW; 64DD7F96FB8C4BC5 CRC64;

Query Match 41.2%; Score 571.5; DB 11; Length 320;
Best Local Similarity 46.1%; Pred. No. 6.6e-37;
Matches 123; Conservative 48; Mismatches 89; Indels 7; Gaps 3;

QY 11 VGFVAGRMAGATAGLIRAGKVEAQHILASAPTDRNLCHFQAL---GCRTHSNQEVQLQ 67
DB 3 VGFAGQALACALARGFTAAAGLSAHKIIASSP-DMDLPTVSALRMGMVNLTRSNKDTVR 61
QY 68 SCLLVIFATKPHVLPVLAELAVPVVTEHILYSVAAGVSLTLEELL---PNTNRVLRVL 124
DB 62 HSDVLFVAVKPHIIPFILDIGADVQARHIVVSCAAGVTISSVEKKLMAFQAPKVICM 121
QY 125 PNLPCVQBGAIVMARGHRVGSSETKLQHLLEACRCCEVPEAYVDIHTGLSGGVAFV 184
DB 122 TNPVVRGATVYATGTHALVEDGKLLBQLMSSVGFCTEVEDLIDAVTGLSGGPAYA 181
QY 185 CAFSEALARGAVKMGMPSSLAHRIAATLLGTAKMLLHGGHQAQLRSVDCVTPGGTTIYG 244
DB 182 FMALDALADGVKMGVPRRLAQLGAALIGAALMLLSDHDPGQLKDNVCSPPGGATIHA 241
QY 245 LHALEGGGLRAATMSAVEAATCRAKEL 271
DB 242 LHFLSGGFRSLINAVEASCIRTREL 268

RESULT 9
Q9VEJ3 PRELIMINARY; PRT; 273 AA.
AC Q9VEJ3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG5840 protein (rb62767p).
GN CG5840
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
```

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo K., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Fan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhan M., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Patel A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Pirela S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celisner S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003718; AAP55428.1; -;
DR EMBL; AY071558; AAL49180.1; -;
DR FlyBase; FBgn0038516; CG5840.
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR TIGRFAMs; TIGR00112; proC; 1.
DR PROSITE; PS00521; P5CR; 1.
SQ SEQUENCE 273 AA; 28099 MW; 563754C62ACE77AB CRC64;

Query Match 40.4%; Score 560.5; DB 5; Length 273;
Best Local Similarity 44.2%; Pred. No. 4e-36;
Matches 118; Conservative 50; Mismatches 96; Indels 3; Gaps 3;

QY 10 RVGFVAGRWAGATAGLIRAGKVEAQHILASAP--PTDR-NLCHFOALGCRTHSHNOEVLQ 67
Db 6 KIGFLGGNNMAKALAKGLAAGLAKPNTLIASVHPADKLSQFSGLGVETVTKNAPVWQ 65
QY 68 SCLVIFATPKPHVLPVAVPVTTEHILSVAGVSLSTLEELLPPNTRVLRVLPNL 127
Db 66 QSDVVFVSVKPVVPSVLSLQIP-LSSGKFLFSVAMGILTSTISSLSPQARVIRVMPNL 124
QY 128 PCVVOEGAIYVARGRHVGSSETKLLQHLLEACRCCEVPAYVDIHTGLSGSVAFVCF 187
Db 125 PAVYVSCSVFVRSKATDADITQKLSQVSTCEPVDSEQLDVTALSGSPAYVFM 184
QY 188 SEALAEAGVKGMPSSLAHRTAQTLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYGLHA 247
Db 185 IEALDAGVHGMFPRDLAYRLASQTVLGAGHVMVDSGMHPQLKDGVTSPAGSTAALRQ 244
QY 248 LEQGLRAATMSAVEAATCRAKELSRK 274
Db 245 LELSGFRAAVSGAVEQATLCRCQISGK 271

RESULT 10

Q922W5

AC Q922W5;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Similar to pyrroline-5-carboxylate reductase 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006727; AAH06727.1; -;
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR TIGRFAMs; TIGR00112; proC; 1.
DR PROSITE; PS00521; P5CR; UNKNOWN 1.
SQ SEQUENCE 309 AA; 32373 MW; A3CD24AACDD53DEF CRC64;
Query Match 40.4%; Score 560.5; DB 11; Length 309;
Best Local Similarity 43.2%; Pred. No. 4.6e-36;
Matches 115; Conservative 54; Mismatches 92; Indels 5; Gaps 2;

QY 11 VGFVAGRWAGATAGLIRAGKVEAQHILASAP--TDRNLCHFOALGCRTHSHNOEVLQS 68
Db 3 VGFIGAGQDAFLAKGFTAGVLAHKIMASSPDDQATVSALRKIGVNLTPHNKETVRH 62
QY 69 CLVIFATPKPHVLPVAVPVTTEHILSVAGVSLSTLEELL---PPNTRVLRVLP 125
Db 63 SDVLFVAVPHIIPFILDIGANIEDRHIVVSCAAGVTINSTEKKLTAFQPAKPVIRWT 122
QY 126 NLPVVOEGAIYVARGRHVGSSETKLLQHLLEACRCCEVPAYVDIHTGLSGSVAFVC 185
Db 123 NTPVVVREGVTYATGTHAQVEDGRLEQMLGMSVGFCTEVEDLIDAVTGLSGSPAYAF 182
QY 186 AFSEALAEAGVKGMPSSLAHRTAQTLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYGL 245
Db 183 TALDALADGGVKGMLPRRLVRLAGLQAALLGAARKMLLDSEQHPSQLKDNVCSPPGATIH 242
QY 246 HALEQGLRAATMSAVEAATCRAKEL 271
Db 243 HVLESGGFRSLINAVEASCIRTREL 268

RESULT 11
Q9HH99
ID Q9HH99 PRELIMINARY; PRT; 270 AA.
AC Q9HH99;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE proC (Pyrroline-5-carboxylate reductase).
GN PROC OR MA4102.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A;
RA Zhang J.K., White A.K., Kuettner H.C., Boccazzi P., Metcalf W.W.;
RT "Molecular and genetic analysis of proline biosynthesis in the
methanogenic archaeon *Methanosarcina acetivorans* C2A.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,

Db 184 VEMADAANLVNGWTRKAYKFAAQSVLGAACKMILETGEHPGKLDKDDVCSGTTIEAVYA 243
Qy 248 LEQGLRAATMSAVEAATCRAKELSRK 274
||: ||: :||:| :|:| |
Db 244 LEKSGFRASVIAAVDACIKKSKLMSK 270

Search completed: July 21, 2003, 09:55:13
Job time : 83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:43:01 ; Search time 23 Seconds
(without alignments)
494.109 Million cell updates/sec

Title: US-09-806-536A-14

Perfect score: 1386

Sequence: 1 MAAEPPSPRRVGVGAGRMA.....AATMSAVEATCRAKELSRK 274

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	584.5	42.2	319	1 PROC_HUMAN	P32322 homo sapien
2	502.5	36.3	273	1 PROC_PEA	Q04708 pisum sativ
3	498	35.9	278	1 PROC_ACTCH	O04016 actinidia c
4	492	35.5	274	1 PROC_SOYBN	P17817 glycine max
5	475	34.3	276	1 PROC_ARATH	P54904 arabidopsis
6	423.5	30.6	269	1 PROC_ECOLI	P00373 escherichia
7	412.5	29.8	267	1 PROC_SYNY3	P74572 synechocyst
8	390.5	28.2	263	1 PROC_TREPA	P27771 treponema p
9	387	27.9	261	1 PROC_TRETH	P54893 thermus the
10	383.5	27.7	294	1 PROC_MYCLE	P46725 mycobacteri
11	378	27.3	297	1 PROH_BACSU	P14383 bacillus su
12	366	26.4	295	1 PROC_MYCTU	Q11141 mycobacteri
13	362.5	26.2	270	1 PROC_CORGL	P46340 corynebacte
14	362	26.1	299	1 PROC_CAEEL	Q20848 caenorhabdi
15	345.5	24.9	272	1 PROC_PSEAE	P22008 pseudomonas
16	344	24.8	282	1 PROC_SCHPO	Q9P7V7 schizosacch
17	337.5	24.4	278	1 PROI_BACSU	P54552 bacillus su
18	321	23.2	286	1 PROC_YEAST	P32263 saccharomyc
19	318.5	23.0	285	1 PROC_AQUAE	O66553 aquifex aeo
20	310.5	22.4	311	1 PROC_NEUCR	Q12641 neurospora
21	307	22.2	278	1 PROC_VIBAL	P52053 vibrio algi
22	305.5	22.0	320	1 PROC_ZALAR	Q12740 zalerion ar
23	301.5	21.8	271	1 PROC_HAEIN	P43869 haemophilus
24	198	14.3	272	1 PROC_BACSU	Q00777 bacillus su
25	189.5	13.7	273	1 CMEA_BACSU	P39696 bacillus su
26	164	11.8	257	1 PROC_HELPU	Q9ZK56 helicobacte
27	159	11.5	257	1 PROC_HELPJ	O25773 helicobacte
28	125.5	9.1	251	1 PROC_METSM	P22350 methanobrev
29	114	8.2	294	1 GARR_ECOLI	P23523 escherichia
30	99.5	7.2	388	1 ALR_MYCAV	Q91888 mycobacteri
31	97	7.0	1503	1 MRP6_HUMAN	O95255 homo sapien
32	96	6.9	401	1 CSD_PSEAE	Q9HXX3 pseudomonas
33	96	6.9	528	1 SERA_MYCTU	O53243 mycobacteri

34 95 6.9 4393 1 PGBM_HUMAN P98160 homo sapien
35 94.5 6.8 388 1 ALR_MYCLE P38056 mycobacteri
36 93.5 6.7 349 1 GPDA_MYCLE Q9CB99 mycobacteri
37 93 6.7 528 1 SERA_MYCLE O33116 mycobacteri
38 92.5 6.7 2436 1 ABC2_HUMAN Q9BZC7 homo sapien
39 92 6.7 278 1 HBD_DEIRA Q9RVG1 deinococcus
40 92 6.6 386 1 ALR_MYCTU Q50705 mycobacteri
41 92 6.6 400 1 NTRB_AZOB R P45670 azospirillum
42 92 6.6 482 1 MURC_PASMU P57818 pasteurella
43 90 6.5 334 1 GPDA_MYCTU P95113 mycobacteri
44 90 6.5 388 1 ALR3_RHILO Q981H7 rhizobium l
45 89.5 6.5 2301 1 POLG_TMEVD P13899 t genome po

ALIGNMENTS

RESULT 1

ID PROC_HUMAN STANDARD; PRT; 319 AA.
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
GN PYCR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112821; PubMed=1730675;
RA Doughterty K.M., Brandriss M.C., Valle D.;
RT "Cloning human pyroline-5-carboxylate reductase cDNA by
complementation in Saccharomyces cerevisiae.";
RL J. Biol. Chem. 267:871-875(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = L-pyrroline-5-
carboxylate + NAD(P)H.
CC -!- PATHWAY: Proline biosynthesis; third (last) step.
CC -!- SUBUNIT: HOMODECAMER OR HOMODECAMER.
CC -!- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; M78836; AAA36407.1; -;
CC EMBL; BC001504; AAH01504.1; -;
CC PIR; A41770; A41770.
CC Genew; HGNC:9721; PYCR1.
CC MIM; 179035; -;
CC InterPro; IPR000304; P5CR.
CC Pfam; PF01089; P5CR; 1.
CC TIGRFAMs; TIGR00112; PROC; 1.
CC PROSITE; PS00521; P5CR; 1.
KW Oxidoreductase; Proline biosynthesis; NADP.
FT CONFLICT 155 155 T -> S (IN REF. 2).
SQ SEQUENCE 319 AA; 33374 MW; F5E74B5BDFB475EF CRC64;

Query Match 42.2%; Score 584.5; DB 1; Length 319;
Best Local Similarity 46.4%; Pred. No. 2.9e+38;
Matches 124; Conservative 47; Mismatches 89; Indels 7; Gaps 3;


```
Db 117 D-WAGHSRIRVMPNTPAAYGEAASVMSLGATGEGELITLFLGAIGKIKWADEKLEF 175
Oy 172 IHTGLSGVAFVCAFEALAEAGVAKMGMPSSLAHRIAQTLLGTAKMLLHGEQHQAOLR 231
Db 176 AVTGLSGGPAYIFLATEALADGVAGAGLPRLEALGLASQTVLGAASVMVARGKHGQQLK 235
Oy 232 SDVCTPGGTTTYGLHALEOGLRAATMSAVEAATCRAKELSRK 274
Db 236 DDVASAGGTTIAGIHELEKGGFRGTLMLNAVVSATKRSQEIFKR 278

RESULT 4
PROC_SOYBN
ID PROC_SOYBN STANDARD; PRT; 274 AA.
AC P17817;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90340278; PubMed=2199815;
RA Delauney A.J., Verna D.P.S.;
RT "A soybean gene encoding delta 1-pyrroline-5-carboxylate reductase
RT was isolated by functional complementation in Escherichia coli and is
RT found to be osmoregulated.";
RL Mol. Gen. Genet. 221:299-305(1990).
CC -I- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = 1-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -I- PATHWAY: Proline biosynthesis; third (last) step.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES, BUT MOSTLY IN
CC NODULES.
CC -I- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X16352; CAA34401.1; -.
CC PIR; S10186; S10186.
CC InterPro; IPR000304; P5CR.
CC Pfam; PF01089; P5CR; 1.
CC TIGRFAMs; TIGR00112; PROC; 1.
CC PROSITE; PS00521; P5CR; 1.
CC Oxidoreductase; Proline biosynthesis; NADP.
SQ SEQUENCE 274 AA; 28586 MW; 933CFCD7598B63 CRC64;

Query Match 35.58; Score 492; DB 1; Length 274;
Best Local Similarity 39.68; Pred. No. 3.4e-31;
Matches 107; Conservative 51; Mismatches 96; Indels 16; Gaps 3;

Oy 11 VGFVAGRMAGATAOGLIRAGKVEAQAHLASAPTDRLNCH-----FOALGCRUTHSN 62
Db 13 LGFIGACKMAESITARGAVSGVLPSPRI-----RTAVHFNLRGAFESFGVTVLP 65
Oy 63 QEVQLSCLLVIFATKPHVLAIEVAPVVTTEHILVSVAAAGVSLSTLEELLPPNTRVLR 122
Db 66 DDVVRESDDVVVLSVKPOLVDVYVKLTPLTKHLLVSVAAAGTKKLDQE-WAGNDRFIR 124
Oy 123 VLPNLPVQVGEALVYVARGHVSSTETKLLQHLLEACGRCEEVPEVYVYDHTGLSGGVA 182
Db 125 VMPTNPAVGOAASVMSLGSATGEDGNIITIAQLFGSIGKTIWKAEEKYFDAITGLSGGPA 184
```

```
Oy 183 FYCAFSEALAECAVKGMPSSLAHRIAQTLLGTAKMLLHGEQHQAOLRSDVCTPGGTTI 242
Db 185 VYVLAIEALADGVAGAGLPRDLUSLASQTVLGAASVMSQTKHGEQQLKDDVTSPEGTTI 244
Oy 243 YGLHALEOGLRAATMSAVEAATCRAKELS 272
Db 245 TGIHELENGGFRGTLMLNAVVAAKRSRELS 274

RESULT 5
PROC_ARATH
ID PROC_ARATH STANDARD; PRT; 276 AA.
AC P54904;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
GN PROC1 OR AT5G14800 OR T9L3.100.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94294559; PubMed=8022935;
RA Verbruggen N., Villarroel R., van Montagu M.;
RT "Osmoregulation of a pyroline-5-carboxylate reductase gene in
RT Arabidopsis thaliana.";
RL Plant Physiol. 103:771-781(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta;
RA Verbruggen N., Villarroel R., Hua X., van Montagu M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Kohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volkkaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entlan K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA Langanh K.-A., McCullagh B., Robben J., Grynoprez B., Zimmermann W.,
RA Rampersger U., Wedler H., Balke K., Wedler E., Peters S.,
RA van Staveren M., Dirke W., Mooljman P., Klein Lankhorst R.,
RA Weitzenger T., Borge G., Rose M., Hauf J., Berner S., Hempel S.,
RA Feldpausch M., Lamberth S., Villarroel R., Giesen J., Ardiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer R.F.X., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:823-826(2000).
CC -I- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = 1-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -I- PATHWAY: Proline biosynthesis; third (last) step.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
```



```
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -|- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = L-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -|- PATHWAY: Proline biosynthesis; third (last) step.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90916; BAA18679.1; -
CC InterPro: IPR000304; P5CR.
CC Pfam: PF01089; P5CR; 1.
CC TIGRfam: TIGR00112; proc; 1.
CC PROSITE: PS00521; P5CR; 1.
CC Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
SQ SEQUENCE 267 AA; 27883 MW; FD9C428047F391B0 CRC64;

Query Match 29.8%; Score 412.5; DB 1; Length 267;
Best Local Similarity 36.2%; Pred. No. 4.6e-25;
Matches 98; Conservative 50; Mismatches 110; Indels 13; Gaps 4;

QY 10 RVGFVAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNL-CHFOALGCRTHSNQEVLSQ 68
DB 4 QLGIIIGGVMAEAILARLIAETKYEPIIIVGEPHGCARDYQVVRSPDQNEANV 63
QY 69 CLLVIFATPKPHVLPVAVLEA- PVTTHILVSAAGVSLSTLELLPPNTRVLRV 123
DB 64 SEVLLAVRPQVLDRLVSLAGGANRPLV- ISILAGVSLQRIQKGFDPDA-IIRA 116
QY 124 LPLPCVQVQGAIVMARGHRVGSSEFKLLQHLLEAGRCCEVEPYAVVDIHTGLSGGVAF 183
DB 117 MPNTPATVAGMTAIANKMVEPDQLAKAKAIFSAVGNVVEPNLMDAVTGVSGSGPAY 176
QY 184 VCAFSALAEAGVAKMGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGTTIY 243
DB 177 VALMIEALADGGVLAGLPRAIAQKALQVLTGTAELIKETEEHPAQIKDKVTSFGTTIA 236
QY 244 GHALBQGGGLRAATNSAVAACTCAKELSRK 274
DB 237 GVAVLEKMGFRSAIEAVRAAYRRSOELGKK 267

RESULT 8
PROC_TREPA STANDARD; PRT; 263 AA.
AC P27771; O83775;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
GN PROC TP00797.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
```

```
RX MEDLINE=90264287; PubMed=2188947;
RA Gherardini F.C., Hobbs M.M., Stamm L.V., Bassford P.J. Jr.;
RT "Complementation of an Escherichia coli proc mutation by a gene cloned
RT from Treponema pallidum.";
RL J. Bacteriol. 172:2996-3002(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -|- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = L-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -|- PATHWAY: Proline biosynthesis; third (last) step.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M73825; AAA27478.1; -
CC PIR; S27734; AAC65760.1; -
CC TIGR; TP0797; -
CC InterPro: IPR000304; P5CR.
CC Pfam: PF01089; P5CR; 1.
CC PROSITE: PS00521; P5CR; 1.
CC Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
CC CONFLICT 1 65
FT FT 77 93 VLRDRQSFQGVKLVLSLA -> GTARSPFQESAISC (IN
FT CONFLICT 256 260 VRAAL -> CRWLS (IN REF. 1).
FT SEQUENCE 263 AA; 27645 MW; ED4AD7C54BAF9D61 CRC64;

Query Match 28.2%; Score 390.5; DB 1; Length 263;
Best Local Similarity 38.1%; Pred. No. 2.3e-23;
Matches 99; Conservative 41; Mismatches 111; Indels 9; Gaps 5;

QY 11 VGFGAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNL-CHFOALGCRTHSNQEVLSQ 69
DB 3 VGFGAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNL-CHFOALGCRTHSNQEVLSQ 62
QY 70 LLVIFATPKPHVLPVAVLEA- PVTTHILVSAAGVSLSTLELLPPN-TRVLRLVP 125
DB 63 EWIFLAVKPSQISTVLRDRQSF--QGKVLISLAAGMSCAAYEALFAADPHQGIHLSLPP 120
QY 126 NLPCVVQGAIVMARGHRH-VGSSETKLLQHLLEAGRCCEVEPYAVVDIHTGLSGGVAFV 184
DB 121 NLPCQVARG-VIIAEARHTLHDEHAALLAVLRTVAQVVDVTFATFAGIACAPAF 179
QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGTTIY 244
DB 180 AQFTEALADAGVRYGLARDQAYRLAAHMLEGTAALIQHSGVHPAQLKDRVCSAGSTIRG 239
QY 245 LHALEQGGGLRAATNSAVEAA 264
```

```
Db 240 VLAEQGLRAVIAVRAA 259
RESULT 9
PROC_THETH
ID PROC_THETH STANDARD; PRT; 261 AA.
AC P34893;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
GN PROC.
OS Thermus thermophilus.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Thermales;
OC Thermaceae; Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB27;
RX MEDLINE=94168609; PubMed=8123043;
RA Hoshino T., Kosuge T., Hidaka Y., Tabata K., Nakahara T.;
RT "Molecular cloning and sequence analysis of the proc gene encoding
RT delta 1-pyrroline-5-carboxylate reductase from an extremely
RT thermophilic eubacterium Thermus thermophilus";
RL Biochem. Biophys. Res. Commun. 199:410-417(1994).
CC -!- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = 1-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -!- PATHWAY: Proline biosynthesis; third (last) step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D25413; BAA05001.1; -
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR PROSITE; PS00521; P5CR; 1.
DR OXidoreductase; Proline biosynthesis; NADP.
DR KW SEQUENCE 261 AA; 27819 MW; 3435C3A8F4299B3E CRC64;
Query Match 27.9%; Score 387; DB 1; Length 261;
Best Local Similarity 37.7%; Pred. No. 4.2e-23;
Matches 101; Conservative 47; Mismatches 102; Indels 18; Gaps 9;
QY 10 RVGFVAGRMAGIAAGLIRAG--KVEAQHILASAPTDRLNCHFOALGCR-TTHSNQEV 66
Db 2 RLAFVGLKMGKRSILKALGERGLRPEEVGLGRTP-ERSRELAEPFGVRLTRADLGM 60
QY 67 QSCLLVIFATKPHVLPVLAIEVAPVVTTEHIL--VSVAGVSLSTLEELLPNTRVL 124
Db 61 ER---VLIIVQPRDFPALAPEA-----HRLGLYSIMAGISTSLARRL-DNRRVVRAM 111
QY 125 PNLPCVVQBGAIWMARGHWGSSSE-TKLQHLLEACGRCCEVPEAYVDIHTGLSGGVAF 183
Db 112 PNLAVVIGESSLTALKAREADLAFARLFAIVGVYVEIPEHLFDFTGMSASAPAY 171
QY 184 VCAFSALAEAGVKMGSPSLAHRITAAOTLLGTAKMLHEGQHQAQLSDVCTPGGTTIY 243
Db 172 LAVVAEALADAGVKMGMPRALRLAALDAALATGELL--KGRHPAQVKDEVASPGGTTIH 229
QY 244 GLHLEQGGLRATNSAVEAATCRAKEL 271
Db 230 GLHLEARAVRAAFYEAVEAATRRGHEL 257
RESULT 10
```

```
Db 183 AYFELLVEALVDAGVAVGLTRQVATELAAQTMAAGSAAMLLRMDQDRHSNAEVLPGAQVD 242
Qy 227 --PAQLRSDVCTPGGTTIYGLHLEQGLRAATMSAVEAATCAKEL 271
    ||||| : ||||| | ||||| : ||||| : ||||| : |||||
Db 243 VPAALRATITSPGGTTAAALRELGRGLRVDDVAAVQAARIRSEOL 289

RESULT 11
PROH_BACSU
ID PROH_BACSU STANDARD; PRT; 297 AA.
AC P14383; O31928; O07508;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase 1 (EC 1.5.1.2) (P5CR 1) (P5C
reductase 1).
GN PROH OR PROC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=1423;
EN [1]
SEQUENCE FROM N.A.
RP
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi N.,
Guiseppi G., Guy B.J., Hega K., Hache J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone E.,
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis."
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE OF 1-256 FROM N.A.
RC STRAIN=W23, and 168;
RX MEDLINE=91192601; PubMed=1849493;
RA Ahn K.S., Wake R.G.;
RT "Variations and coding features of the sequence spanning the
replication terminus of Bacillus subtilis 168 and W23 chromosomes."
RL Gene 98:107-112(1991).
RN [3]
RP SEQUENCE OF 1-200 FROM N.A.
RC STRAIN=W23;
RX MEDLINE=89155440; PubMed=2493444;
RA Lewis P.J., Wake R.G.;
RT "DNA and protein conservation at the replication terminus in
Bacillus subtilis 168 and W23."
RL J. Bacteriol. 171:1402-1408(1989).
RN [4]
RP SEQUENCE OF 199-297 FROM N.A., AND FUNCTION.
ID PROC_MYCTU STANDARD; PRT; 295 AA.
QY 9 RVGVFGAGRMAGATAGLIIRAGKVEAQHILASAPTDRLNCHFOALGCRTHSNOEVL-- 66
Db 17 KKVAFAGSMAGMISGIVRANKIPKONI-----CVTKNSRTERLTE 59
QY 67 -----QSCL-----LVIPATKPHVLPVLAEPVVTTEHILSVAAAGVSLST 109
Db 60 LELOYGIRKALPNQCIEDMDVLLAMKPKDAENALSSLSKSIOPHQLLSVLACITTSF 119
QY 110 LEELPPNTRVRLVLPNLCVVOEGATVNRGHRVGSSETKLQLHLEACGCEVPRAY 169
Db 120 IEQSLLEQPVVVRVMPNTSSMIGASATAIALGKYVSEDLKLAELALGCMGEVYIQENQ 179
QY 170 VDIHTGLSGSVAFVFCAPSEALAEAGVKMGPMSSLAHRTAAOTLLGTAKMLLHGOHPAQ 229
Db 180 MDIFTGIAGSGPAYFYLLMEFTEKTEGEEAGLDKQLSRSIGTQTLGAARKMLMETGEHPEI 239
QY 230 LRSDVCTPGGTTIYGLHLEQGLRAATMSAVEAATCAKELS 272
Db 240 LRDNITSPNGTTAAGLQALKSGGGEATSOAIKHAARKSKEIS 282
RESULT 12
PROC_MYCTU
ID PROC_MYCTU STANDARD; PRT; 295 AA.
QY 9 RVGVFGAGRMAGATAGLIIRAGKVEAQHILASAPTDRLNCHFOALGCRTHSNOEVL-- 66
Db 17 KKVAFAGSMAGMISGIVRANKIPKONI-----CVTKNSRTERLTE 59
QY 67 -----QSCL-----LVIPATKPHVLPVLAEPVVTTEHILSVAAAGVSLST 109
Db 60 LELOYGIRKALPNQCIEDMDVLLAMKPKDAENALSSLSKSIOPHQLLSVLACITTSF 119
QY 110 LEELPPNTRVRLVLPNLCVVOEGATVNRGHRVGSSETKLQLHLEACGCEVPRAY 169
Db 120 IEQSLLEQPVVVRVMPNTSSMIGASATAIALGKYVSEDLKLAELALGCMGEVYIQENQ 179
QY 170 VDIHTGLSGSVAFVFCAPSEALAEAGVKMGPMSSLAHRTAAOTLLGTAKMLLHGOHPAQ 229
Db 180 MDIFTGIAGSGPAYFYLLMEFTEKTEGEEAGLDKQLSRSIGTQTLGAARKMLMETGEHPEI 239
QY 230 LRSDVCTPGGTTIYGLHLEQGLRAATMSAVEAATCAKELS 272
Db 240 LRDNITSPNGTTAAGLQALKSGGGEATSOAIKHAARKSKEIS 282
```



```
QY 67 QSCLLVFATKPHVLPVLAIEAPV---TTEHLVSVAGVSLSTLEELLPTNTRVLRV 123
Db 60 DEADVFLVCKPFIVEVLEISEITGTIDNNNSAOSVVVSMAGISIAAMEESASAGLPVVRV 119
QY 124 LPLPCVVOBGATVMAARGHVSSTKLLQHLLEACGRCEVEPEAYVDIHTGLSGGVAF 183
Db 120 MPNTPLVKGGMSTVTKGRYVDAEQLEQVKDLLSTVGDVLEVAESDIDAVTAMSGSPAY 179
QY 184 VCAFSSEALAEAGVKGMPSSLAHRIAQTLLGTAKMLLHGGQHPAQLRSDVCTPGTITTY 243
Db 180 LFLVTEALIEAGVNLGLPRATAKKLAVASFEGAATMMKETGKPSSELRAVSGVSPAGTTVA 239
QY 244 GLHALBOGGLRAATMSAVEAATCRAKELSRK 274
Db 240 AIRELEESGIRGAFYRAAQACADRSEELGKR 270

RESULT 14
PROC_PSEAE
ID PROC_PSEAE STANDARD; PRT; 299 AA.
AC Q20848;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative pyrroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C
DE reductase).
DE F55G1.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Murray J., Le T.T.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = L-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -!- PATHWAY: Proline biosynthesis; third (last) step.
CC -!- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U58750; AAB00645.1; -.
CC Wormpep; F55G1.9; CE07286.
CC InterPro; IPR00304; P5CR.
CC Pfam; PF01089; P5CR; 1.
CC PROSITE; PS00521; P5CR; 1.
CC Hypothetical protein; X03057 MW; 75D2D6032E0A5670 CRC64;
CC SEQUENCE 299 AA; 32057 MW; 75D2D6032E0A5670 CRC64;

Query Match 26.1%; Score 362; DB 1; Length 299;
Best Local Similarity 33.5%; Pred. No. 4.1e-21;
Matches 93; Conservative 53; Mismatches 114; Indels 18; Gaps 6;

QY 13 FVAGNRMAGIAOGLIRACKVEAQHILASPTDRNLCHQALGCRTHSNQ-EVLQ--SC 69
Db 19 FIGGNNMAAIIKGGKONKQTPKSNIVIGVQTEKSAEKWRLGKYNFTNLEMLERYST 78
QY 70 LLVFATKPHVLPVLAIEAPVVTTEHLVSVAGVSLSTLEELLP---PNTRLVRLPN 126
Db 79 AIYVICVKPVFEVSS-WPVNSRPEFFISVWAGVPLKVLNKLFPVSGNTTIVRLMPN 137
QY 127 LPCVVOBGATVMAARGHVS---GSSETKLQLHLEACGRCEVEPEAYVDIHTGLSGGVA 182
Db 120 MPNTPLVKGGMSTVTKGRYVDAEQLEQVKDLLSTVGDVLEVAESDIDAVTAMSGSPAY 179
```

```
Db 138 VASSIGAGASTMCEKNEKIMNQDSHIELAREFAECVGTVELIPERCNPAMAIGGSSPA 197
QY 183 FVCAFSSEALAEAGVKGMPSSLAHRIAQTLLGTAKMLLH-----EGQHPAQLRSDVC 235
Db 198 WTFMYTIESLADGAVAQGLGRAEKRLAAQAQVLAQVLAQVLAQVLAQVLAQVLAQVLAQV 257
QY 236 TPGGTTIYGLHALBOGGLRAATMSAVEAATCRAKELSR 273
Db 258 SPGGTTIEGVRALEKNGFRYAVMEAVAASTKADEMAK 295

RESULT 15
PROC_PSEAE
ID PROC_PSEAE STANDARD; PRT; 272 AA.
AC P22008;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyrroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
DE PROC OR PA0393.
GN Pseudomonas aeruginosa.
OS Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-12.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=90185238; PubMed=2107123;
RA Savioz A., Jeenes D.J., Kocher H.P., Haas D.;
RT "Comparison of proC and other housekeeping genes of Pseudomonas
RT aeruginosa with their counterparts in Escherichia coli.";
RL Gene 86:107-111(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
RN [3]
RP SEQUENCE OF 1-6 FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=91285432; PubMed=1676385;
RA Whitchurch C.B., Hobbs M., Livingston S.P., Krishnapillai V.,
RA Mattick J.S.;
RT "Characterisation of a Pseudomonas aeruginosa twitching motility gene
RT and evidence for a specialised protein export system widespread in
RT eubacteria.";
RL Gene 101:33-44(1991).
CC -!- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = L-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -!- PATHWAY: Proline biosynthesis; third (last) step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M33557; AAA25975.1; -.
CC EMBL; AB004476; AAG03782.1; -.
CC EMBL; M55524; AAA25958.1; -.
CC
```

